PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/17, 15/18, 5/14, C12Q 1/02, C07K 16/26, A01K 67/027, C07K 14/435	**A1**	(11) International Publication Number: **WO 00/32779** (43) International Publication Date: 8 June 2000 (08.06.00)

(21) International Application Number: PCT/EP99/09316

(22) International Filing Date: 30 November 1999 (30.11.99)

(30) Priority Data:
09/201,226   30 November 1998 (30.11.98)   US

(71)(72) Applicant and Inventor: LEPTIN, Maria [DE/DE]; Heinestrasse 19, D–50931 Köln (DE).

(74) Agent: SCHÜSSLER, Andreas; Patentanwälte Huber & Schüßler, Truderinger Strasse 246, D–81825 München (DE).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NUCLEIC ACIDS AND POLYPEPTIDES OF *D. MELANOGASTER* INSULIN-LIKE GENES AND USES THEREOF
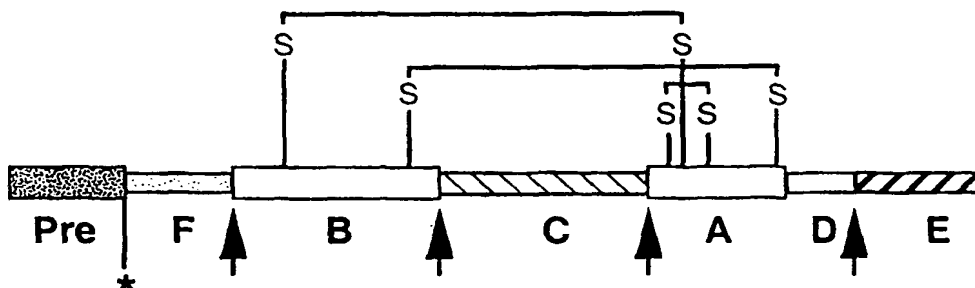
(57) Abstract

The present invention relates to a *D.melanogaster* insulin–like gene and methods for identifying insulin–like genes. The methods provide nucleotide sequences of a *D.melanogaster* insulin–like gene, amino acid sequences of its encoded proteins, and derivatives (e.g., fragments) and analogs thereof. The invention further relates to fragments (and derivatives and analogs thereof) of insulin–like proteins which comprise one or more domains of an insulin–like protein. Antibodies to an insulin–like protein, and derivatives and analogs thereof, are provided. Methods of production of an insulin–like protein (e.g., by recombinant means), and derivatives and analogs thereof, are provided. Further, methods to identify the biological function of a *D.melanogaster* insulin–like gene are provided, including various methods for the functional modification (e.g., overexpression, underexpression, mutation, knock–out). Still further, methods to identify a *D.melanogaster* gene that modifies the function of, and/or functions in a signaling pathway with an insulin–like gene are provided. The invention further provides uses of *Drosophila* insulin–like nucleic acids and proteins, e.g., as media additives, and as pesticides.

NUCLEIC ACIDS AND POLYPEPTIDES OF *D. MELANOGASTER* INSULIN-LIKE GENES AND USES THEREOF

PRIORITY APPLICATION

This application claims priority to U.S. Ser. No. 09/201,226 (Leptin) filed
5   November 30, 1998.

BACKGROUND OF THE INVENTION

Insulin is the central hormone governing metabolism in vertebrates (reviewed in
Steiner et al., 1989, In *Endocrinology*, DeGroot, eds. Philadelphia, Saunders: 1263-1289).
10   In humans, insulin is secreted by the beta cells of the pancreas in response to elevated
blood glucose levels which normally occur following a meal. The immediate effect of
insulin secretion is to induce the uptake of glucose by muscle, adipose tissue, and the liver.
A longer term effect of insulin is to increase the activity of enzymes that synthesize
glycogen in the liver and triglycerides in adipose tissue. Insulin can exert other actions
15   beyond these "classic" metabolic activities, including increasing potassium transport in
muscle, promoting cellular differentiation of adipocytes, increasing renal retention of
sodium, and promoting production of androgens by the ovary. Defects in the secretion
and/or response to insulin are responsible for the disease diabetes mellitus, which is of
enormous economic significance. Within the United States, diabetes mellitus is the fourth
20   most common reason for physician visits by patients; it is the leading cause of end-stage
renal disease, non-traumatic limb amputations, and blindness in individuals of working
age (Warram et al., 1995, In *Joslin's Diabetes Mellitus*, Kahn and Weir, eds., Philadelphia,
Lea & Febiger, pp. 201-215; Kahn et al., 1996, Annu. Rev. Med. 47:509-531; Kahn, 1998,
Cell 92:593-596). Beyond its role in diabetes mellitus, the phenomenon of insulin
25   resistance has been linked to other pathogenic disorders including obesity, ovarian
hyperandrogenism, and hypertension.

Insulin-like proteins are a large and widely-distributed group of structurally-related
peptide hormones that have pivotal roles in controlling animal growth, development,
reproduction, and metabolism. Consequently, the insulin superfamily has become one of
30   the most intensively investigated classes of peptide hormones. Studies of insulin-like

1

molecules in invertebrates have been motivated by the desire to identify proteins that play

analogous roles to the well-characterized activities of insulin and IGF in mammals.

Although insulin superfamily members in invertebrates have been less extensively

analyzed than in vertebrates, a number of different subgroups have been defined. Such

5   subgroups include molluscan insulin-related peptides (MIP-I to MIP-VII) (Smit et al.,

1988, Nature 331:535-538; Smit et al., 1995, Neuroscience 70:589-596), the bombyxins of

lepidoptera (originally referred to as prothoracicotropic hormone or PTTH) (Kondo et al.,

1996, J Mol. Biol. 259:926-937), and the locust insulin-related peptide (LIRP) (Lagueux et

al., 1990, Eur. J. Biochem. 187:249-254). Most recently, there have been descriptions of

10   an exceptionally large insulin-like gene family in the nematode C. elegans

(WO1999US08522; Duret, et al., 1998, Genome Res. 8:348-353; Brousseau, et al., 1998,

Early 1998 East Coast Worm Meeting, abstract 20; Kawano, et al., 1998, Worm Breeder's

Gazette 15(2):47; Pierce and Ruvkun, 1998, Early 1998 East Coast Worm Meeting,

abstract 150; Wisotzkey and Liu, 1998, Early 1998 East Coast Worm Meeting, abstract

15   206). Also, putative orthologs of both vertebrate insulin and IGF have been identified in a

tunicate (McRory and Sherwood, 1997, DNA and Cell Biology 116:939-949). From the

extensive sequence divergence evident among known subfamilies of insulin-like proteins,

it is assumed that this is an ancient family of regulatory hormones that evolved to control

growth, reproduction and metabolism in early metazoans. However, the precise

20   evolutionary origins of this important family remain unclear.

       Early attempts to propagate Drosophila cells in culture revealed a growth factor

requirement in defined medium which could be provided by purified bovine insulin,

implying the existence of a related endogenous factor in. Also, bovine and human insulin

were found to stimulate the differentiation of Drosophila cells grown in culture (Seecof

25   and Dewhurst, 1974, Cell Differ. 3(1):63-70; Pimentel, et al., 1996, Biochem. Biophys.

Res. Commun. 226(3):855-61). One report described the presence of an "insulin-like

activity" in unpurified Drosophila extracts that elicited a hypoglycemic effect when

injected into mice, although the activity was not particularly strong (Meneses and De Los

Angeles Ortiz, 1975, Comp. Biochem. Physiol. A. 51(2):483-5). Another group (LeRoith,

30   et al., 1981, Diabetes 30(1):70-6) fractionated an insulin-like material from Drosophila

based on immunoreactivity and showed that this material had insulin-like activity on

isolated rat adipocytes. Also, polyclonal antibodies raised against bovine/porcine insulin

were used to localize insulin-immunoreactive material in Drosophila tissue (Gorczyca, et

2

al., 1993, J. Neurosci. 13(9):3692-704), and specific insulin-immunoreactive substances were detected at neuromuscular junctions and in the central nervous system. However, these substances were not characterized further to validate that they correspond to *bona fide* insulin proteins at the level of primary protein sequence. Indeed, despite this long

5   history of phenomenological evidence for insulin-like activities, true insulin-like genes and proteins in *Drosophila* have not been identified and characterized at the sequence level.

More compelling evidence for evolutionary conservation of insulin-like signaling pathways in *Drosophila* has come from the identification of an apparent homolog of the insulin receptor (Petruzzelli et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714). One

10   insulin receptor homolog has been characterized thus far in *Drosophila*, termed InR (insulin receptor) also known as DIR (*Drosophila* insulin receptor) (Ruan et al., 1995, J. Biol. Chem. 270:4236-4243), which exhibits extensive homology with vertebrate insulin and IGF receptors in both the extracellular ligand-binding domain and the intracellular tyrosine kinase domain. Genetic analysis of InR function in *Drosophila* has revealed that

15   it is an essential gene with an apparent role in the development of the epidermis and nervous system, as well as growth control (Fernandez et al., 1995, EMBO J. 14:3373-3384). Flies that are homozygous for mutations in InR generally exhibit an embryonic lethal phenotype, but flies bearing certain heteroallelic combinations of InR mutations live to adulthood and the surviving animals have about 50% the normal body

20   weight (Garafalo, Chen, et al., 1996, Endocrinology 137(3):846-56). This result is reminiscent of a similar phenotype observed in mutant mice lacking functional IGF-I receptor genes (Liu, et al., 1993, Cell 75(1):59-72). Aside from this potential role of InR in growth regulation, the role, if any, that InR may have in metabolic regulation in *Drosophila* remains unclear. The ligand binding specificity of InR has been examined

25   using in vitro assays for receptor activation/phosphorylation, and competitive binding of test ligands compared to porcine insulin (Fernandez-Almonacid and Rosen, 1987, Mol. Cell Biol. 7(8):2718-27). Curiously, the results of this study indicated that InR binds vertebrate insulin, and does not apparently recognize vertebrate IGF-I or IGF-II, or even bombyxin-II from the silkworm, implying that the natural *Drosophila* ligands for InR may

30   bear more structural resemblance to vertebrate insulin than to these other insulin superfamily proteins.

The structural homologies of components of the Drosophila *InR* pathway with those involved in insulin signaling in mammals, as well as the function of the *InR* pathway

3

in controlling ﹘ ﹍wth, and the circumstantial evidence for *Drosophila* insulin-like

activities, raise critical questions with respect to further analysis of this pathway and its

potential applications. Important issues regarding the biological function, regulation, and

signaling mechanisms of insulin superfamily hormones could best be addressed if these

5   pathways could be analyzed using model genetic organisms. In particular, the facile

genetic tools currently available in two model organisms, the fruit fly *Drosophila*

melanogaster and the nematode *Caenorhabditis elegans*, have proven to be of enormous

utility in defining the biological function of genes through mutational analysis, as well as

for identifying the components of biochemical pathways conserved during evolution with

10   large-scale, systematic genetic screens (Scangos, 1997, Nature Biotechnol. 15:1220-1221;

Miklos and Rubin, 1996, Cell 86:521-529). Key discoveries regarding constituents of a

number of important human disease pathways, such as the Ras pathway and the pathway

controlling programmed cell death, first came from genetic analysis of pathways known to

have an evolutionary relation in *Drosophila* and *C. elegans*, and later shown to have direct

15   relevance to human biology (Yuan et al., 1993, Cell 75:641-652; Therrien et al., 1995, Cell

83:879-888; Karim et al., 1996, Genetics 143:315-329; Kornfeld et al., 1995, Cell

83:903-913; Rubin et al., 1997, "Protein kinase required for Ras signal transduction", U.S.

Patent No. 5,700,675; Steller et al., 1997, "Cell death genes of *Drosophila melanogaster*

and vertebrate homologs", U.S. Patent No. 5,593,879).

20

SUMMARY OF THE INVENTION

The present invention relates to proteins encoded by nucleotide sequences of a

D. melanogaster insulin-like gene, as well as fragments and other derivatives and

analogs of such insulin-like proteins. Nucleic acids encoding the insulin-like gene and

25   fragments or derivatives are also within the scope of the invention. Production of the

foregoing proteins, *e.g.*, by recombinant methods, is provided.

The invention also relates to insulin-like protein derivatives and analogs which

are functionally active, *i.e.*, which are capable of displaying one or more known

functional activities associated with a full-length (wild-type) insulin-like protein.

30   Examples of such functional activities include antigenicity (ability to bind, or to

compete for binding, to an anti-insulin-like protein antibody), immunogenicity (ability

to generate antibody which binds to an insulin-like protein), and ability to bind (or

4

compete for bi...uing) to a receptor for insulin (*e.g.*, that is encoded by the *D. melanogaster* insulin receptor-like gene, *InR*).

The invention further relates to fragments (and derivatives and analogs thereof) of an insulin-like protein which comprise one or more domains of the insulin-like protein.

5   Antibodies to an insulin-like protein, its derivatives and analogs, are additionally provided.

Methods for genetic analysis of pathways involving insulin superfamily hormones in *Drosophila* are provided. Such methods may yield results of importance to human disease. For example, systematic identification of participants in intracellular signaling by

10   insulin-like hormones, or components regulating secretion and turnover of insulin-like hormones, provide leads to the identification of drug targets, therapeutic proteins, diagnostics, or prognostics useful for treatment or management of insulin resistance in diabetics.

15

BRIEF DESCRIPTION OF FIGURES

FIG. 1 illustrates the structural organization of precursor forms of the insulin superfamily of hormones. The different domains that make up precursor forms of insulin-like hormones are represented as boxes labeled Pre, F, B, C, A, D, and E.

20   Domains that may remain in a mature hormone are represented as unshaded boxes (the B, A, and D peptide domains) or as lightly hatched (the C or "connecting" peptide domain). Domains that are removed during proteolytic processing are represented as shaded (the Pre peptide domain and F domain) or as hatched (the E peptide domain). IGF hormones are unique in having D and E peptide domains; these domains are represented as smaller

25   boxes in FIG. 1. Cleavage sites utilized by proteases during proteolytic processing (*i.e.*, protein maturation) are indicated below the boxes. The asterisk marks the position of cleavage by signal peptidase. Arrows indicate cleavage sites by prohormone convertases. Disulfide bonds (S-S) are represented above the boxes with lines indicating connections between covalently-bonded Cys residues.

30   FIG. 2 illustrates conserved structural features of known insulin superfamily members. The aligned sequences of the B and A chain peptide domains are shown for representative insulin superfamily hormones from the following vertebrates and invertebrates: human insulin (Bell et al., 1979, Nature 29:525-527), human IGF-I (Jansen

5

et al., 1983, Nature 306:609-611), human relaxin 1 (Hudson et al., 1983, Nature
301:628-631), RLF from human (Adham al., 1993, J. Biol. Chem. 268:26668-26672),
placentin from human (Chassin et al., 1995, Genomics 29:465-470), bombyxin II from
silkworm (Nagasawa et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83:5840-5843), MIP from
5   freshwater snail (Smit et al., 1988, Nature 331:535-538), and LIRP from locust (Lagueux
et al., 1990, Eur. J. Biochem. 187:249-254). The numbering scheme shown at the bottom
of the figure is for residues of the A and B chains relative to residue numbers for human
insulin peptide domains. The nearly invariant positions of the six Cys residues that
participate in disulfide bonds are boxed. MIP-I is unusual in having two extra Cys
10   residues which are also individually boxed in that sequence. Other conserved amino acid
positions that play important roles in promoting the common insulin superfamily fold are
highlighted by shading of the following residue positions: B6, B8, B11, B15, B18, A2,
A16, and A19. Three helical regions that comprise the common insulin fold are marked
above the alignments using a "<--->" symbol.
15   FIG. 3 shows key structural features for *D. melanogaster* Insulin-like protein
folding and conserved Cysteine residues in vertebrate superfamily. Numbers shown in
parentheses represents the number of residues omitted from the C peptide sequence.

          FIG. 4 shows annotated DNA sequence of dIns1 insulin-like gene. Nucleic acid
sequence of dIns1 is set forth in SEQ ID NO:1. Amino acid sequence of dIns1 is set forth
20   in SEQ ID NO:2.

DETAILED DESCRIPTION OF THE INVENTION

          Described herein is a novel insulin-like gene from *Drosophila* and the
characterization of its function. The *Drosophila* insulin-like gene described herein is
25   much closer in structure to vertebrate insulins than the insulin-like proteins found in the
nematode *C. elegans*. This new insulin-like gene in *Drosophila* constitutes a very useful
tool for probing the function and regulation of its corresponding pathways. Systematic
genetic analysis of signaling pathways involving insulin-like proteins in *Drosophila* can be
expected to lead to the discovery of new drug targets, therapeutic proteins, diagnostics and
30   prognostics useful in the treatment of diseases and clinical problems associated with the
function of insulin superfamily hormones in humans and other animals, as well as clinical
problems associated with aging and senescence. Furthermore, analysis of these same
pathways using *Drosophila* insulin-like proteins as tools will have utility for identification

and validation of pesticide targets in invertebrate pests that are components of these signaling pathways.

     Use of a *Drosophila* insulin-like gene for such purposes as disclosed herein, has advantages over manipulation of other known components of the fruit fly InR pathway

5   including *InR, Pi3K92E,* and *chico.* First, use of ligand-encoding a *Drosophila* insulin-like gene provides a superior approach for identifying factors that are upstream of the receptor in the signal transduction pathway. Specifically, components involved in the synthesis, activation and turnover of insulin-like proteins may be identified. Furthermore, the discovery of multiple, different insulin-like hormones provides a rational approach to

10   separate components involved in responses to different, specific environmental or regulatory signals. This is less technically feasible with manipulation of downstream components of the pathway found in target tissues. Further, the diversity of different insulin-like hormones provides a means to identify potential new receptor and/or signal transduction systems for insulin superfamily hormones that are structurally different from

15   those that have been characterized to date, in either vertebrates or invertebrates. Still further, use of *Drosophila* as a system for analyzing the function and regulation of an insulin-like gene has great advantages over approaches in other organisms due to the ability to rapidly carry out large-scale, systematic genetic screens as well as the ability to screen small molecules directly on whole organisms for possible therapeutic or pesticide

20   use. Particularly, the *Drosophila* insulin-like gene described herein is significantly closer in structure to vertebrate insulin hormones than the insulin-like proteins of *C. elegans*; therefore, the fruit fly *Drosophila* may serve as a better model for vertebrate insulin function and signaling than the nematode *C. elegans* due to this greater structural similarity. Moreover, the fruit fly *Drosophila* is clearly the preferred genetic model

25   organism for dissecting the function of insulin-like proteins, and validating potential pesticide targets, with respect to other insect pest species.

Isolation Of a *D. Melanogaster* Insulin-Like Gene

     In specific embodiments, insulin-like nucleic acids of the invention comprise the

30   cDNA sequence of SEQ ID NO:1, or the coding regions thereof, or nucleic acids encoding an insulin-like protein (*e.g.*, a protein having the sequence of SEQ ID NO:2). As used herein, a gene "corresponding" to a cDNA sequence shall be construed to mean the gene that encodes the RNA from which the cDNA is derived. The invention provides purified

or isolated nuc...c acids consisting of at least 8 nucleotides (*i.e.*, a hybridizable portion) of an insulin-like gene sequence; in other embodiments, the nucleic acids consist of at least 25 (continuous) nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200 nucleotides of an insulin-like sequence, or a full-length insulin-like coding sequence. In

5   another embodiment, the nucleic acids are smaller than 35, 200, or 500 nucleotides in length. Nucleic acids can be single or double stranded. The invention also relates to nucleic acids hybridizable to or complementary to the foregoing sequences or their reverse complements. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of

10 an insulin-like gene.

In the above or alternative embodiments, the nucleic acids of the invention consist of a nucleotide sequence of not more than 2, 5, 10, 15, or 20 kilobases.

Hybridization Conditions

15 A nucleic acid which is hybridizable to an insulin-like nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NO:1, or to its reverse complement, or to a nucleic acid encoding an insulin-like derivative, or to its reverse complement), under conditions of high, medium, or low stringency is provided. Methods for selection of appropriate conditions for such stringencies is well known in the art (*see e.g.*, Sambrook et al., 1989,

20 *supra*; Ausubel et al., eds., in the Current Protocols in Molecular Biology series of laboratory technique manuals, 8 1987-1997, Current Protocols, 8 1994-1997 John Wiley and Sons, Inc.).

An example of suitable conditions of high stringency that can be used is as follows. Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in

25 buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 g/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture containing 100 g/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is

30 followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography.

An example of procedures using conditions of medium stringency is as follows. Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll,

8

1% BSA, and 5ʋ0 g/ml denatured salmon sperm DNA. Hybridizations are carried out in
the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2%
BSA, 100 g/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm
^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at

5   40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM
Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh
solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for
autoradiography. If necessary, filters are washed for a third time at 65-68°C and
re-exposed to film.

10   Conditions of low stringency are as follows. Incubation for 8 hours to overnight at
37°C in a solution comprising 20% formamide, 5x SSC, 50mM sodium phosphate (pH
7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared
salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of
filters in 1x SSC at 37°C for 1 hour.

15   Fragments of insulin-like nucleic acids comprising regions conserved between (*i.e.*,
with homology to) other insulin-like nucleic acids, of the same or different species, are
also provided. Nucleic acids encoding one or more insulin-like domains are provided.

Cloning Procedures

20   The insulin-like gene of the invention can be cloned using any suitable technique
known in the art (*see* Sambrook *et al.* (1989), *supra*; DNA Cloning: A Practical Approach,
Vol. 1, 2, 3, 4, (1995) Glover, ed., MRL Press, Ltd., Oxford, U.K.). For example, with
expression cloning, an expression library is constructed, mRNA is isolated, and cDNA is
made and ligated into an expression vector (*e.g.*, a bacteriophage derivative) that it is

25   capable of being expressed by the host cell into which it is then introduced. Various
screening assays can then be used to select for the expressed insulin-like product such as
immunoassays using anti-insulin-like antibodies.

Polymerase chain reaction (PCR) can used to amplify the desired sequence in a
genomic or cDNA library, prior to selection. Oligonucleotide primers representing known

30   insulin-like sequences can be used as primers in PCR. Preferably, the oligonucleotide
primers represent at least part of conserved segments of strong homology between
insulin-like genes of different species. The synthetic oligonucleotides may be utilized as
primers to amplify sequences from a source (RNA or DNA), preferably a cDNA library, of

9

potential intere.... PCR can be carried out, e.g., by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (e.g., Gene AmpJ). The nucleic acid being amplified can include mRNA or cDNA or genomic DNA from any species. One may synthesize degenerate primers for amplifying homologs from other species in the PCR reactions. The

5   stringency of hybridization conditions used in priming the PCR reactions can be varied to allow for greater or lesser degrees of nucleotide sequence similarity between the known insulin-like nucleotide sequences and a nucleic acid homolog (or ortholog) being isolated. For cross species hybridization, low stringency conditions are preferred. For same species hybridization, moderately stringent conditions are preferred. After successful

10   amplification of a segment of an insulin-like homolog, that segment may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone. This, in turn, permits the determination of the gene's complete nucleotide sequence, the analysis of its expression, and the production of its protein product for functional analysis, as described below. In this fashion, additional genes encoding

15   insulin-like proteins and insulin-like analogs may be identified.

Any eukaryotic cell potentially can serve as the nucleic acid source for molecular cloning of an insulin-like gene. The nucleic acid sequences encoding insulin-like proteins may be isolated from vertebrate, mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, insects (e.g., Drosophila),

20   invertebrates, plants, etc. The DNA may be obtained by standard procedures known in the art from cloned DNA (e.g., a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at

25   specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques such as agarose and polyacrylamide gel electrophoresis and column chromatography.

30   Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if a portion of an insulin-like gene or its specific RNA or a fragment thereof is available and can be purified and labeled, the generated DNA fragments may be screened

10

by nucleic acid nybridization to the labeled probe (Benton and Davis, 1977, Science 196:180; Grunstein and Hogness, 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of

5   fragment sizes with those expected according to a known restriction map if such is available. Further selection can be carried out on the basis of the properties of the gene.

Alternatively, the presence of the desired gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be

10   selected and expressed to produce a protein that has, e.g., similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, hormonal activity, binding activity, or antigenic properties as known for an insulin-like protein. Using an antibody to a known insulin-like protein, other insulin-like proteins may be identified by binding of the labeled antibody to expressed putative insulin-like proteins,

15   e.g., in an ELISA (enzyme-linked immunosorbent assay)-type procedure. Further, using a binding protein specific to a known insulin-like protein, other insulin-like proteins may be identified by binding to such a protein (see e.g., Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898).

An insulin-like gene can also be identified by mRNA selection using nucleic acid

20   hybridization followed by in vitro translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified insulin-like DNA of another species (e.g., Drosophila, mouse, human). Immunoprecipitation analysis or functional assays (e.g., aggregation ability in vitro, binding to receptor, etc.) of the in vitro translation products of the isolated products of the

25   isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against insulin-like protein. A radiolabeled insulin-like cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabeled

30   mRNA or cDNA may then be used as a probe to identify the insulin-like DNA fragments from among other genomic DNA fragments.

Alternatives to isolating the insulin-like genomic DNA include, chemically synthesizing the gene sequence itself from a known sequence or making cDNA to the

11

mRNA which encodes the insulin-like protein. For example, RNA for cDNA cloning of the insulin-like gene can be isolated from cells which express the gene.

The identified and isolated gene can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible

5   vectors include, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Suitable vectors include bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene USA, La Jolla, California). The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has

10   complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction

15   endonuclease recognition sequences. In an alternative method, the cleaved vector and an insulin-like gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, etc., so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after

20   insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired gene, for example, by size fractionization, can be done before insertion into the cloning vector.

In an additional embodiment, the desired gene may be identified and isolated after insertion into a suitable cloning vector using a strategy that combines a "shot gun"

25   approach with a "directed sequencing" approach. Here, for example, the entire DNA sequence of a specific region of the genome, such as a sequence tagged site (STS), can be obtained using clones that molecularly map in and around the region of interest.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate an isolated insulin-like gene, cDNA, or synthesized DNA

30   sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

12

The insulin-like sequences provided by the instant invention include those nucleotide sequences encoding substantially the same amino acid sequences as found in native insulin-like proteins, and those encoded amino acid sequences with functionally equivalent amino acids, as well as those encoding other insulin-like derivatives or analogs,

5   as described below for insulin-like derivatives and analogs.

Expression Of a _D. Melanogaster_ Insulin-Like Gene

The nucleotide sequence coding for an insulin-like protein or a functionally active analog or fragment or other derivative thereof, can be inserted into any

10   appropriate expression vector that contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native insulin-like gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence such as mammalian cell systems infected with virus (_e.g._,

15   vaccinia virus, adenovirus, _etc._); insect cell systems infected with virus (_e.g._, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation

20   elements may be used. In yet another embodiment, a fragment of an insulin-like protein comprising one or more domains of the insulin-like protein is expressed.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein

25   coding sequences. These methods may include _in vitro_ recombinant DNA and synthetic techniques and _in vivo_ recombinants (genetic recombination). Expression of a nucleic acid sequence encoding an insulin-like protein or peptide fragment may be regulated by a second nucleic acid sequence so that the insulin-like protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of an

30   insulin-like protein may be controlled by any promoter/enhancer element known in the art including those of prokaryotic expression vectors and plant expression vectors; promoter elements from yeast or other fungi; and transcriptional control regions. In some embodiments, the promoter will exhibit tissue specificity. In a specific

13

embodiment, a vector is used that comprises a promoter operably linked to an insulin-like gene nucleic acid, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

Expression constructs can be made by subcloning an insulin-like coding

5   sequence into the EcoRI restriction site of each of the three pGEX vectors (Smith and Johnson, 1988, Gene 7:31-40). This allows for the expression of the insulin-like protein product from the subclone in the correct reading frame. Expression vectors containing insulin-like gene inserts can be identified by three general approaches: (a) nucleic acid hybridization; (b) presence or absence of "marker" gene functions; and (c) expression of

10   inserted sequences. In the first approach, the presence of an insulin-like gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted insulin-like gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine

15   kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, *etc.*) caused by the insertion of an insulin-like gene in the vector. For example, if the insulin-like gene is inserted within the marker gene sequence of the vector, recombinants containing the insulin-like insert can be identified by the absence of the marker gene function. In the third approach, recombinant

20   expression vectors can be identified by assaying the insulin-like product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the insulin-like protein in *in vitro* assay systems, *e.g.*, binding with anti-insulin-like protein antibody.

Once a particular recombinant DNA molecule is identified and isolated, several

25   methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. Some of the expression vectors which can be used include human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda phage), and plasmid and

30   cosmid DNA vectors.

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain

14

inducers; thus, expression of the genetically engineered insulin-like protein may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g., glycosylation, phosphorylation of proteins. Appropriate cell lines or host systems

5   can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce a non-glycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may

10   effect processing reactions to different extents.

     In other embodiments of the invention, the insulin-like protein, fragment, analog, or derivative may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence of a different protein). Such a chimeric product can be

15   made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, e.g., by use of a peptide synthesizer.

20

Identification And Purification Of Gene Products

     In particular aspects, the invention provides amino acid sequences of insulin-like proteins and fragments and derivatives thereof which comprise an antigenic determinant (i.e., can be recognized by an antibody) or which are otherwise functionally active, as well

25   as nucleic acid sequences encoding the foregoing. "Functionally active" insulin-like material as used herein refers to that material displaying one or more functional activities associated with a full-length (wild-type) insulin-like protein, e.g., binding to an insulin-like receptor (e.g., InR or insulin-like protein binding partner, antigenicity (binding to an anti- insulin-like protein antibody), immunogenicity, etc.

30   In specific embodiments, the invention provides fragments of an insulin-like protein consisting of at least 10 amino acids, 20 amino acids, 50 amino acids, or of at least 75 amino acids. In other embodiments, the proteins comprise or consist essentially of an insulin-like B peptide domain, an insulin-like A peptide domain, an insulin-like C peptide

15

domain, or any combination of the foregoing, of an insulin-like protein. Fragments, or proteins comprising fragments, lacking some or all of the foregoing regions of a insulin-like protein are also provided. Nucleic acids encoding the foregoing are provided. In specific embodiments, the foregoing proteins or fragments are not more than 25, 50, or

5   100 contiguous amino acids.

Once a recombinant which expresses the insulin-like gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labeling of the product followed by analysis by gel electrophoresis, immunoassay, *etc.*

10   Once the insulin-like protein is identified, it may be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay.

15   Alternatively, once an insulin-like protein produced by a recombinant is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, *see* Hunkapiller et al., 1984, Nature 310:105-111).

20   In another alternate embodiment, native insulin-like proteins can be purified from natural sources, by standard methods such as those described above (*e.g.*, immunoaffinity purification).

In a specific embodiment of the present invention, such insulin-like proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods or

25   by purification of native proteins, include but are not limited to those containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in FIG. 4 (SEQ ID NO:2), as well as fragments and other derivatives, and analogs thereof, including proteins homologous thereto.

30   **Structure Of Insulin-Like Genes And Proteins**

The structure of insulin-like genes and proteins of the invention can be analyzed by various methods known in the art, including genetic analysis and protein analysis.

Genetic analysis methods for determining the structure of cloned DNA or cDNA

16

corresponding to an insulin-like include Southern hybridization, Northern hybridization, restriction endonuclease mapping, and DNA sequence analysis. Accordingly, this invention provides nucleic acid probes recognizing an insulin-like gene. For example, polymerase chain reaction followed by Southern hybridization with an insulin-like

5   gene-specific probe can allow the detection of an insulin-like gene in DNA from various cell types. Methods of amplification other than PCR are commonly known and can also be employed. In one embodiment, Southern hybridization can be used to determine the genetic linkage of an insulin-like gene. Northern hybridization analysis can be used to determine the expression of an insulin-like gene. Various cell types, at various states of

10   development or activity can be tested for insulin-like gene expression. The stringency of the hybridization conditions for both Southern and Northern hybridization can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific insulin-like gene probe used. Modifications of these methods and other methods commonly known in the art can be used.

15   Restriction endonuclease mapping can be used to roughly determine the genetic structure of an insulin-like gene. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis.

        DNA sequence analysis can be performed by any techniques known in the art, such as the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the

20   Sanger dideoxy method (Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), or use of an automated DNA sequenator (e.g., Applied Biosystems, Foster City, California).

        The amino acid sequence of an insulin-like protein can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the protein, e.g., with

25   an automated amino acid sequencer. An insulin-like protein sequence can be further characterized by a hydrophilicity analysis (Hopp and Woods, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the insulin-like protein and the corresponding regions of the gene sequence that encode such regions.

30   Secondary, structural analysis (Chou and Fasman, 1974, Biochemistry 13:222) can also be done, to identify regions of an insulin-like protein that assume specific secondary structures.

        Manipulation, translation, and secondary structure prediction, open reading

17

frame prediction and plotting, as well as determination of sequence homologies, can also be accomplished using computer software programs available in the art.

Other methods of structural analysis include X-ray crystallography, nuclear magnetic resonance spectroscopy and computer modeling.

5

Antibodies

The insulin-like protein of SEQ ID NO:2, or fragments or derivatives thereof, may be used as an immunogen to generate antibodies. Such antibodies include polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In 10 another embodiment, antibodies to a domain (e.g., an insulin-like receptor binding domain) of an insulin-like protein are produced. In a specific embodiment, fragments of an insulin-like protein identified as hydrophilic are used as immunogens for antibody production using art-known methods. Some examples of suitable techniques include methods which provides for the production of antibody molecules by continuous cell lines 15 in culture; the production of monoclonal antibodies in germ-free animals (see e.g., PCT/US90/02545); the use of human hybridomas (Cole et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030); transforming human B cells with EBV virus in vitro (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). Additionally, known techniques can be used for the production of "chimeric antibodies" 20 (e.g. by splicing the genes from a mouse antibody molecule specific for an insulin-like protein together with genes from a human antibody molecule of appropriate biological activity), insulin-like-specific single chain antibodies; and Fab expression libraries (e.g. to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for insulin-like proteins, derivatives, or analogs). The foregoing antibodies can 25 be used against the insulin-like protein sequences described herein, e.g., for imaging these proteins, measuring levels thereof, in diagnostic methods, etc.

Insulin-Like Proteins, Derivatives And Analogs

The invention relates to insulin-like proteins and derivatives, fragments, and 30 analogs thereof, as well as the nucleic acids encoding them. In one embodiment, the insulin-like proteins are encoded by the insulin-like nucleic acids described above. In particular aspects, the proteins, derivatives, or analogs are of insulin-like proteins encoded by the sequence of SEQ ID NO:1.

18

In a specific embodiment, the insulin-like protein fragment, derivative or analog is functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type insulin-like protein. As one example, such fragments, derivatives or analogs have the desired immunogenicity or antigenicity for use in

5   immunoassays, for immunization, for inhibition of insulin-like activity, *etc.* As another example, such fragments, derivatives or analogs which have the desired binding activity can be used for binding to the InR gene product. As yet another example, they have the desired binding activity can be used for binding to a binding protein specific for a known insulin-like protein (*see e.g.*, Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et

10   al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898). Derivatives or analogs that retain, or alternatively lack or inhibit, a desired insulin-like protein property-of-interest (*e.g.*, binding to an insulin-like protein binding partner), can be used as inducers, or inhibitors, respectively, of such property and its physiological correlates. A specific embodiment relates to an insulin-like protein fragment that can be bound by an anti-insulin-like protein

15   antibody. Derivatives or analogs of an insulin-like protein can be tested for the desired activity by procedures known in the art.

In particular, insulin-like derivatives can be made by altering insulin-like sequences by substitutions, additions (*e.g.*, insertions) or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences,

20   other DNA sequences which encode substantially the same amino acid sequence as an insulin-like gene may be used in the practice of the present invention. These include nucleotide sequences comprising all or portions of an insulin-like gene which is altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the insulin-like

25   derivatives of the invention include those containing, as a primary amino acid sequence, all or part of the amino acid sequence of an insulin-like protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity

30   which acts as a functional equivalent, resulting in a silent alteration. Substitutions for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine.

19

The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Such substitutions are generally understood to be conservative

5   substitutions.

In a specific embodiment of the invention, proteins consisting of or comprising a fragment of an insulin-like protein of at least 10 (continuous) amino acids of the insulin-like protein is provided. In other embodiments, the fragment comprises at least 20, 50, or 75 amino acids of the insulin-like protein. In specific embodiments, such

10 fragments are not larger than 35, 100 or 200 amino acids. Derivatives or analogs of insulin-like proteins may comprise regions that are substantially homologous to an insulin-like protein or fragment thereof (e.g., in various embodiments, at least 60% or 70% or 80% or 90% or 95% identity over an amino acid sequence of identical size. As used herein, "percent (%) sequence identity" with respect to a subject sequence, or a

15 specified portion of a subject sequence, is defined as the percentage of nucleotides or amino acids in the candidate sequence identical with the nucleotides in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul et al., J. Mol. Biol. (1997) 215:403-410;

20 http://blast.wustl.edu/blast/README.html; (hereinafter referred to generally as "BLAST") with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A percent (%) identity value is

25 determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported.

The insulin-like derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, a cloned insulin-like gene sequence can

30 be modified by any of numerous strategies known in the art (Sambrook et al., 1989). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated in vitro. Additionally, an insulin-like nucleic acid sequence can be mutated in vitro or in vivo, to create and/or

20

destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, for example, chemical mutagenesis, *in vitro* site-directed mutagenesis

5   (Hutchinson et al., 1978, J. Biol. Chem. 253:6551), use of TAB7 linkers (Pharmacia), PCR with primers containing a mutation, *etc.*

Manipulations of an insulin-like protein sequence may also be made at the protein level. Included within the scope of the invention are insulin-like protein fragments or other derivatives or analogs which are differentially modified during or after translation,

10   *e.g.,* by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, *etc.* Any of numerous chemical modifications may be carried out by known techniques, such as specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction,

15   metabolic synthesis in the presence of tunicamycin, *etc.*

In addition, analogs and derivatives of an insulin-like protein can be chemically synthesized. For example, a peptide corresponding to a portion of an insulin-like protein which comprises the desired domain, or which mediates the desired activity *in vitro,* can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical

20   amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the insulin-like sequence. Examples of non-classical amino acids include the D-isomers of the common amino acids, -amino isobutyric acid, 4- aminobutyric acid, Abu, 2-amino butyric acid, -Abu, -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine,

25   citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, -alanine, fluoro-amino acids, designer amino acids such as -methyl amino acids, C-methyl amino acids, N-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In a specific embodiment, an insulin-like protein derivative is a chimeric or fusion

30   protein comprising an insulin-like protein or fragment thereof (preferably consisting of at least a domain or motif of the insulin-like protein, or at least 10 amino acids of the insulin-like protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. In specific embodiments, the amino acid

21

sequence of the different protein is at least 6, 10, 20 or 30 continuous amino acids of the different proteins or a portion of the different protein that is functionally active. In one embodiment, such a chimeric protein is produced by recombinant expression of a nucleic acid encoding the protein (comprising an insulin-like-coding sequence joined in-frame to a

5   coding sequence for a different protein). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, e.g., by use of a peptide synthesizer.

10 Chimeric genes comprising portions of an insulin-like gene fused to any heterologous protein-encoding sequences may be constructed. A specific embodiment relates to a chimeric protein comprising a fragment of an insulin-like protein of at least six amino acids, or a fragment that displays one or more functional activities of the insulin-like protein.

15 In another specific embodiment, the insulin-like derivative is a molecule comprising a region of homology with a insulin-like protein. By way of example, in various embodiments, a first protein region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95% identical, when compared to any sequence in

20 the second region of an equal number of amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a molecule can comprise one or more regions homologous to an insulin-like domain or a portion thereof.

In a specific embodiment, the invention relates to insulin-like derivatives and

25 analogs, in particular insulin-like fragments and derivatives of such fragments, that comprise, or alternatively consist of, one or more domains of an insulin-like protein, including but not limited to an insulin-like B peptide domain, an insulin-like A peptide domain, or an insulin-like connecting (C) peptide domain.

A specific embodiment relates to molecules comprising specific fragments of an

30 insulin-like protein that are those fragments in the respective insulin-like proteins of the invention most homologous to specific fragments of a human or mouse insulin-like prot in. A fragment comprising a domain of an insulin-like homolog can be identified by protein analysis methods well known in the art. In another specific embodiment, a

22

molecule is provided that comprises one or more domains (or functional portion thereof) of an insulin-like protein. In particular examples, insulin-like protein derivatives are provided that contain either an A peptide domain or a B peptide domain. By way of another example, such a protein may retain such domains separated by a peptide spacer.

5   Such spacer may be the same as or different from an insulin-like connecting (C) peptide. In another embodiment, a molecule is provided that comprises one or more domains (or functional portion(s) thereof) of an insulin-like protein, and that has one or more mutant (e.g., due to deletion or point mutation(s)) domains of an insulin-like protein (e.g., such that the mutant domain has decreased function).

10

Generation And Genetic Analysis Of _Drosophila_ With Altered Insulin-Like Genes

The present invention provides for methods of creating genetically-engineered fruit flies and laboratory-generated mutant fruit flies.

Genetically-engineered fruit flies can be made that harbor one or more deletions or
15   insertions in an insulin-like gene or genes. In another embodiment, genetically-engineered fruit flies harbor interfering RNAs derived from such genes. In another embodiment, genetically-engineered fruit flies harbor one or more transgenes for mis-expression of wild-type or mutant forms of such genes. Mutant fruit flies can be generated in the lab to contain deletions, insertions, rearrangements, or point mutations in an insulin-like gene or
20   genes, or combinations thereof.

The present invention provides a method by which _Drosophila_ strains with laboratory-generated alterations in insulin-like genes may be used for the identification of insulin-like genes that participate in particular biochemical and/or genetic pathways. In a specific embodiment, _Drosophila_ strains with laboratory-generated alterations in one or
25   more insulin-like genes may be used for the identification of insulin-like genes that participate in biochemical and/or genetic pathways that constitute possible pesticide targets, as judged by phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction. That is, development of such a phenotype in a _Drosophila_ containing an alteration in a _Drosophila_ insulin-like
30   gene indicates that the insulin-like gene is a potential pesticide target.

In another embodiment, _Drosophila_ strains with laboratory-generated alterations relate to therapeutic applications associated with the insulin superfamily hormones, such as metab lic control, growth regulation, differentiation, reproduction, and aging.

23

In another embodiment, *Drosophila* strains with laboratory-generated alterations relate to large-scale genetic modifier screens aimed at systematic identification of components of genetic and/or biochemical pathways that serve as novel drug targets, diagnostics, prognostics, therapeutic proteins, pesticide targets or protein pesticides.

5   The invention provides methods for creating and analyzing *Drosophila* strains having modified expression of insulin-like genes, as described below. In one embodiment, expression modification methods include any method known to one skilled in the art. Specific examples include chemical mutagenesis, transposon mutagenesis, antisense RNA interference, and transgene-mediated mis-expression. In the creation of transgenic

10   animals, it is preferred that heterologous (*i.e.*, non- native) promoters be used to drive transgene expression.

Generation Of Loss-Of-Function Mutation In Insulin-Like Gene

The present invention provides methods of testing for preexisting mutations in a *D.*

15   *melanogaster* insulin-like gene. Mutations in genes that map to the same genetic region as the insulin-like gene (chromosomal band 67C-D) are of particular interest. For example, a large number of previously identified mutations have been mapped to the approximate genetic region of the insulin (67C-D), including l(3)67BDa, l(3)67BDb, l(3)67BDc, l(3)67BDd, l(3)67BDe, l(3)67BDf, l(3)67BDg l(3)67BDh, l(3)67BDi l(3)67BDj,

20   l(3)67BDk, l(3)67BDl, l(3)67BDm, l(3)67BDn, l(3)67BDp, l(3)67BDq, l(3)67BDr (FlyBase: a *Drosophila* database, Flybase consortium, Harvard University); however, the normal function of these genes has not been determined. To ascertain whether any of these mutations are in an insulin-like gene, a genomic fragment containing the *Drosophila* insulin gene and potential flanking regulatory regions can be subcloned into any

25   appropriate *Drosophila* transformation vector, such as the Carnegie series of vectors (Rubin and Spradling, 1983, Nucleic Acids Res. 11(18):6341-51), the pCaspeR series of vectors (Thummel, et al., 1988, Gene 74(2):445-56), or the pW8 vector (Klemenz, et al., 1987, Nucleic Acids Res. 15(10):3947-59) and injected into flies along with an appropriate helper plasmid to supply transposase. Resulting transformants are crossed for

30   complementation testing to an existing panel of *Drosophila* lines containing mutations that have been mapped to the appropriate genomic region (67C-D) as described above (Greenspan, 1997, in *Fly pushing: The Theory and Practice of Drosophila Genetics* Cold Spring Harbor Press, Plainview, NY, pp. 3-46). If a mutant line is discovered to be

24

rescued by this genomic fragment, as judged by complementation of the mutant
phenotype, progressively smaller subclones or clones containing a single insulin gene can
be individually tested until the responsible locus is identified.

5   **Generating Loss-Of-Function Mutations By Mutagenesis**

Further, the invention herein provides a method for generating loss-of-function
mutations in a *D. melanogaster* insulin-like gene. Mutations can be generated by one of
many mutagenesis methods known to investigators skilled in the art (Ashburner, 1989, In
Drosophila: A Laboratory Manual, Cold Spring Harbor, NY, Cold Spring Harbor
10   Laboratory Press: pp. 299-418.; *"Fly pushing: The Theory and Practice of Drosophila
Genetics"* Cold Spring Harbor Press, Plainview, NY). In a specific embodiment, the
mutagens that can be used include but are not restricted to: transposons such as the P or
hobo elements; chemical mutagens such as ethylmethane sulfonate (EMS), methylmethane
sulfonate (MMS), N-ethyl-N-nitrosourea (ENU), triethylmelamine, diepoxyalkanes,
15   ICR-170, or formaldehyde; and irradiation with X-rays, gamma rays, or ultraviolet
radiation.

Mutagenesis by P elements, or marked P elements, is particularly appropriate for
isolation of loss-of-function mutations in *Drosophila* insulin-like genes due to the precise
molecular mapping of these genes, the small size of these targets, the availability and
20   proximity of preexisting P element insertions for use as a localized transposon source, and
the potential to knock out several of these genes by induction of a small deletion of the
locus (Hamilton and Zinn, 1994, Methods in Cell Biology 44:81-94; Wolfner and
Goldberg, 1994, Methods in Cell Biology 44:33-80; Clark, et al., 1994, Proc. Natl. Acad.
Sci. U.S.A. 91(2):719-22; Kaiser, 1990, Bioessays 12(6):297-301, In *Drosophila
25   melanogaster: Practical Uses in Cell and Molecular Biology*, L.S.B. Goldstein and E. A.
Fyrberg, Eds., Academic Press, Inc. San Diego, California). For the purposes of
mutagenesis, modified P elements are typically used which contain one or more of the
following elements: sequences encoding a dominant visible marker, usually a wild-type
white+ or *rosy+* eye color gene, to allow detection of animals containing the P element and
30   to screen for transposition events (Rubin and Spradling, 1982, Science 218(4570):348-53;
Klemenz, et al., 1987, Nucleic Acids Res. 15(10):3947-59), bacterial plasmid sequences
including a selectable marker such as ampicillin resistance to facilitate cloning of genomic
sequences adjacent to the insertion site (Steller and Pirrotta, 1985, Embo. J. 4:167-171)

25

and *lacZ* sequences fused to a weak general promoter to detect the presence of enhancers with a developmental expression pattern of interest (Bellen, et al., 1989, Genes Dev. 3(9):1288-300; Bier, et al., 1989, Genes Dev. 3(9):1273-87; Wilson, et al., 1989, Genes Dev. 3(9):1301-13). For examples of marked P elements useful for mutagenesis *see*

5   "FlyBase - A Drosophila Database", Nucleic Acids Research 26:85-88, (http://flybase.bio.indiana.edu).

A preferred method of transposon mutagenesis employs the "local hopping" method (Tower et al., 1993, Genetics 133:347-359). Briefly, an existing mutant *Drosophila* line containing a P element inserted into chromosomal bands 67C-D, such as

10   l(3)01859 or any other P element that maps within this region, is crossed to a *Drosophila* line expressing transposase in order to mobilize the transposon. Transposition of the P element, which contains a marker gene that typically affects eye color, is determined phenotypically on the basis of eye color change in the resulting progeny. Candidate insertion lines are selected for further analysis on the basis of close linkage of the new

15   insertion to the initial insertion site, which can be determined by standard genetic mapping techniques such as high frequency cosegregation of markers. Each new P insertion line can be tested molecularly for transposition of the P element into the insulin-like gene by assays based on PCR amplification. For each reaction, one PCR primer is used that is homologous to sequences contained within the P element and a second primer is

20   homologous to one of the individual insulin genes, in either the coding region or flanking regions of the insulin-like gene. Products of the PCR reactions are detected by agarose gel electrophoresis. The sizes of the resulting DNA fragments are used to map the site of P element insertion.

Alternatively, Southern blotting and restriction mapping using DNA probes

25   derived from genomic DNA or cDNAs of the insulin-like gene can be used to detect transposition events that rearrange the genomic DNA of the insulin-like gene. P transposition events that map to the insulin gene can be assessed for phenotypic effects in heterozygous or homozygous mutant *Drosophila*, as described in detail below.

30   **Generating Loss-Of-Function Phenotypes Using Methods Based On RNA-Mediated Interference With Gene Expression**

The invention further provides a method for generating loss-of-function phenotypes using methods based on RNA-mediated interference with gene expression.

26

The function of the *Drosophila* insulin-like gene identified herein may be characterized and/or determined by generating loss-of-function phenotypes through such RNA-based methods.

In one embodiment, loss-of-function phenotypes are generated by antisense RNA
5   methods (Schubiger and Edgar, 1994, Methods in Cell Biology 44:697-713). One form of the antisense RNA method involves the injection of embryos with an antisense RNA that is partially homologous to the gene-of-interest (in this case an insulin-like gene). Another form of the antisense RNA method involves expression of an antisense RNA partially homologous to the gene-of-interest by operably joining a portion of the gene-of-interest in
10   the antisense orientation to a powerful promoter that can drive the expression of large quantities of antisense RNA, either generally throughout the animal or in specific tissues. Examples of powerful promoters that can be used in this strategy of antisense RNA include heat shock gene promoters or promoters controlled by potent exogenous transcription factors, such as GAL4 and tTA, described in more detail in the following
15   section. Antisense RNA-generated loss-of-function phenotypes have been reported previously for several *Drosophila* genes including *cactus*, *pecanex*, and *Krupple* (LaBonne, et al., 1989, Dev. Biol. 136(1):1-16; Schuh and Jackle, 1989, Genome 31(1):422-5; Geisler, et al., 1992, Cell 71(4):613-21).

In a second embodiment, loss-of-function phenotypes are generated by
20   cosuppression methods (Bingham, 1997, Cell 90(3):385-7; Smyth, 1997, Curr. Biol. 7(12):793-5; Que and Jorgensen, 1998, Dev. Genet. 22(1):100-9). Cosuppression is a phenomenon of reduced gene expression produced by expression or injection of a sense strand RNA corresponding to a partial segment of the gene-of-interest. Cosuppression effects have been employed extensively in plants to generate loss-of-function phenotypes,
25   and there is report of cosuppression in *Drosophila* where reduced expression of the *Adh* gene was induced from a *white-Adh* transgene (Pal-Bhadra, et al., 1997, Cell 90(3):479-90).

In a third embodiment, loss-of-function phenotypes may be generated by double-stranded RNA interference. This method is based on the interfering properties of
30   double-stranded RNA derived from the coding regions of genes. Termed dsRNAi, this method has proven to be of great utility in genetic studies of the nematode *C. elegans* (*see* Fire et al., 1998, Nature 391:806-811). In a preferred embodiment of this method, complementary sense and antisense RNAs derived from a substantial portion of a

27

gene-of-interest, such as an insulin-like gene, are synthesized *in vitro*. Phagemid DNA templates containing cDNA clones of the gene-of-interest are inserted between opposing promoters for T3 and T7 phage RNA polymerases. Alternatively, one can use PCR products amplified from coding regions of insulin-like gene, where the primers used for

5   the PCR reactions are modified by the addition of phage T3 and T7 promoters. The resulting sense and antisense RNAs are annealed in an injection buffer, and the double-stranded RNA injected or otherwise introduced into animals. Progeny of the injected animals are then inspected for phenotypes-of-interest.

10 ## ANTISENSE REGULATION OF GENE EXPRESSION

The invention provides for antisense uses of a *D. melanogaster* insulin-like gene. In a specific embodiment, an insulin-like protein function is inhibited by use of insulin-like antisense nucleic acids. The present invention provides for use of nucleic acids of at least six nucleotides that are antisense to a gene or cDNA encoding an

15 insulin-like protein or a portion thereof. An insulin-like "antisense" nucleic acid as used herein refers to a nucleic acid capable of hybridizing to a sequence-specific (i.e. non-poly A) portion of an insulin-like RNA (preferably mRNA) by virtue of some sequence complementarily. Antisense nucleic acids may also be referred to as inverse complement nucleic acids. The antisense nucleic acid may be complementary to a coding and/or

20 noncoding region of an insulin-like mRNA. Such antisense nucleic acids have utility in inhibiting an insulin-like protein function. For example, such antisense nucleic acids may be useful as pesticides to eradicate parasites in plants, or in animals such as dogs, horses, and cattle.

The antisense nucleic acids of the invention can be oligonucleotides that are

25 double-stranded or single-stranded, RNA or DNA or a modification or derivative thereof, which can be directly administered to a cell, or which can be produced intracellularly by transcription of exogenous introduced sequences. In a preferred embodiment, the antisense nucleic acids of the invention are double-stranded RNA mentioned previously (*see* Fire et al., 1998, Nature 391:806-811).

30 The insulin-like antisense nucleic acids of the invention are preferably oligonucleotides (ranging from 6 to about 50 oligonucleotides). In specific aspects, an oligonucleotide is at least 10 nucleotides, at least 15 nucleotides, at least 100 nucleotides, or at least 200 nucleotides in length. The oligonucleotide can be DNA or RNA or

28

chimeric mixtures or derivatives or modified versions thereof, or single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, or agents facilitating transport across the cell membrane (see e.g., Letsinger et

5   al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. U.S.A. 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or the blood-brain barrier (see e.g., PCT Publication No. WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (see e.g., Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents (see e.g., Zon, 1988, Pharm. Res.

10   5:539-549).

In a preferred aspect of the invention, an insulin-like antisense oligonucleotide is provided as single-stranded DNA. In another preferred aspect, such an oligonucleotide comprises a sequence antisense to the sequence encoding a B peptide domain or an A peptide domain of an insulin-like protein. The oligonucleotide may be modified at any

15   position on its structure with substituents generally known in the art.

The insulin-like antisense oligonucleotide may comprise at least one modified base moiety for example, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil,

20   dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5.-methoxycarboxymethyluracil, 5-methoxyuracil,

25   2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. In another embodiment, the oligonucleotide comprises at least one

30   modified sugar moiety, for example, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a

29

methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the oligonucleotide is an -anomeric oligonucleotide. An -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual -units, the strands run parallel to each other (Gautier

5   et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide may be conjugated to another molecule, e.g., a peptide, a hybridization-triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially

10   available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (Stein et al., 1988, Nucl. Acids Res. 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

15   In a specific embodiment, an insulin-like antisense oligonucleotide comprises catalytic RNA, or a ribozyme (see e.g., PCT Publication WO 90/11364, published October 4, 1990; Sarver et al., 1990, Science 247:1222-1225). In another embodiment, the oligonucleotide is a 2.-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett.

20   215:327-330).

In an alternative embodiment, the insulin-like antisense nucleic acid of the invention is produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced in vivo such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid

25   (RNA) of the invention. Such a vector would contain a sequence encoding the insulin-like antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and

30   expression in mammalian cells. Expression of the sequence encoding the insulin-like antisense RNA can be by any promoter known in the art. Such promoters can be inducible or constitutive. Examples include the SV40 early promoter region (Benoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3. long terminal repeat of Rous

sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), etc.

5   The antisense nucleic acids of the invention comprise a sequence complementary to at least a sequence-specific portion of an RNA transcript of an insulin-like gene. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable

10   duplex; in the case of double-stranded insulin-like antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an insulin-like RNA it may contain and still form a stable duplex (or

15   triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine, e.g., the melting point of the hybridized complex.

Generating Gain-Of-Function Phenotypes By Ectopic Expression Of An Insulin-Like
20 **Gene**

           The current invention provides methods for generating gain-of-function phenotypes by ectopic expression of an insulin-like gene. Ectopic expression, including mis-expression or overexpression, of a wild type or altered *Drosophila* insulin-like gene in transgenic animals is another useful method for the analysis of gene function (Brand, et al.,

25   1994, Methods in Cell Biology 44:635-654, Ectopic expression in *Drosophila*; Hay, et al., 1997, Proc. Natl. Acad. Sci. U.S.A. 94(10):5195-200). Such transgenic *Drosophila* may be created that contain gene fusions of the coding regions of the insulin-like gene (from either genomic DNA or cDNA) operably joined to a specific promoter and transcriptional enhancer whose regulation has preferably been well characterized, preferably heterologous

30   promoters/enhancers that do not normally drive the expression of the insulin-like gene. Examples of promoters/enhancers that can be used to drive such misexpression of the insulin-like gene include the heat shock promoters/enhancers from the *hsp70* and *hsp83* genes, useful for temperature induced expression; tissue specific promoters/enhancers such

31

as the *sevenless* promoter/enhancer (Bowtell, et al., 1988, Genes Dev. 2(6):620-34), the *eyeless* promoter/enhancer (Bowtell, et al., 1991, Proc. Natl. Acad. Sci. U.S.A. 88(15):6853-7), and *glass*-responsive promoters/enhancers (Quiring, et al., 1994, Science 265:785-9) useful for expression in the eye; enhancers/promoters derived from the *dpp* or

5 *vetigal* genes useful for expression in the wing (Staehling-Hampton, et al., 1994, Cell Growth Differ. 5(6):585-93; Kim, et al., 1996, Nature 382:133-8) and binary control systems employing exogenous DNA regulatory elements and exogenous transcriptional activator proteins, useful for testing the misexpression of genes in a wide variety of developmental stage-specific and tissue-specific patterns. Two examples of binary

10 exogenous regulatory systems include the UAS/GAL4 system from yeast (Hay, et al., 1997, Proc. Natl. Acad. Sci. U.S.A. 94(10):5195-200; Ellis, et al., 1993, Development 119(3):855-65) and the "Tet system" derived from *E. coli*, which are described below. It is readily apparent to those skilled in the art that additional binary systems can be used which are based on other sets of exogenous transcriptional activators and cognate DNA

15 regulatory elements in a manner similar to that for the UAS/GAL4 system and the Tet system.

In a specific embodiment, the UAS/GAL4 system is used. This system is a well-established and powerful method of mis-expression in *Drosophila* which employs the UAS$_G$ upstream regulatory sequence for control of promoters by the yeast GAL4

20 transcriptional activator protein (Brand and Perrimon, 1993, Development 118(2):401-15). In this approach, transgenic *Drosophila*, termed "target" lines, are generated where the gene-of-interest (e.g. an insulin-like gene) to be mis-expressed is operably fused to an appropriate promoter controlled by UAS$_G$. Other transgenic *Drosophila* strains, termed "driver" lines, are generated where the GAL4 coding region is operably fused to

25 promoters/enhancers that direct the expression of the GAL4 activator protein in specific tissues, such as the eye, wing, nervous system, gut, or musculature. The gene-of-interest is not expressed in the so-called target lines for lack of a transcriptional activator to "drive" transcription from the promoter joined to the gene-of-interest. However, when the UAS-target line is crossed with a GAL4 driver line, mis-expression of the gene-of-interest

30 is induced in resulting progeny in a specific pattern that is characteristic for that GAL4 line. The technical simplicity of this approach makes it possible to sample the effects of directed mis-expression of the gene-of-interest in a wide variety of tissues by generating one transgenic target line with the gene-of-interest, and crossing that target line with a

32

panel of pre-existing driver lines. A very large number of specific GAL4 driver lines have been generated previously and are available for use with this system.

In a second embodiment, a related method of directed mis-expression in *Drosophila* is used, that makes use of a tetracycline-regulated gene expression from *E.*
5   *coli*, referred to as the "Tet system". In this case, transgenic *Drosophila* driver lines are generated where the coding region for a tetracycline-controlled transcriptional activator (tTA) is operably fused to promoters/enhancers that direct the expression of tTA in a tissue-specific and/or developmental stage-specific manner. Also, transgenic *Drosophila* target lines are generated where the coding region for the gene-of-interest to be
10   mis-expressed (e.g. an insulin-like gene) is operably fused to a promoter that possesses a tTA-responsive regulatory element. Here again, mis-expression of the gene-of-interest can be induced in progeny from a cross of the target line with any driver line of interest; moreover, the use of the Tet system as a binary control mechanism allows for an additional level of tight control in the resulting progeny of this cross. When *Drosophila* food is
15   supplemented with a sufficient amount of tetracycline, it completely blocks expression of the gene-of-interest in the resulting progeny. Expression of the gene-of-interest can be induced at will simply by removal of tetracycline from the food. Also, the level of expression of the gene-of-interest can be adjusted by varying the level of tetracycline in the food. Thus, the use of the Tet system as a binary control mechanism for mis-expression
20   has the advantage of providing a means to control the amplitude and timing of mis-expression of the gene-of-interest, in addition to spatial control. Consequently, if a gene-of-interest (e.g. an insulin-like gene) has lethal or deleterious effects when mis-expressed at an early stage in development, such as the embryonic or larval stages, the function of the gene-of-interest in the adult can still be assessed using the Tet system, by
25   adding tetracycline to the food during early stages of development and removing tetracycline later so as to induce mis-expression only at the adult stage.

Analysis Of Mutant Phenotypes

After isolation of fruit flies carrying a mutated or mis-expressed insulin-like gene,
30   or inhibitory RNAs, animals are carefully examined for phenotypes-of-interest. For the situations involving deletions, insertions, point mutations, or mis-expression of the insulin-like gene, fruit flies are generated that are homozygous and heterozygous for the altered insulin-like gene.

33

Examples of specific phenotypes that may be investigated include: altered body shape, altered body size, lethality, sterility, reduced brood size, increased brood size, altered life span, defective locomotion, altered body plan, altered cell size, increased cell division, decreased cell division, altered feeding, slowed development, increased

5   development, altered metabolism, (such as altered glycogen synthesis, storage, or degradation; altered lipid synthesis, storage or degradation; altered levels of carbohydrate in the hemolymph; and altered levels of lipid in the hemolymph), and altered morphogenesis of specific organs and tissues such as gonad, nervous system, fat body, hemocytes, peripheral sensory organs, bristles, imaginal discs, eye, wing, leg, antennae,

10   gut, or musculature. For example, it is of particular interest to identify the ligand or ligands responsible for activating InR (or DIR), a *Drosophila* homologue of the insulin receptor. A likely phenotype of a loss-of-function mutation in the ligand for the InR receptor might resemble one or more of the identified loss of function phenotypes for the receptor itself, including reduced body size and weight, reduced female fertility, increased

15   developmental time, and/or defective embryonic neurogenesis.

Methods for creation and analysis of transgenic *Drosophila* strains having modified expression of genes are well known to those skilled in the art (Brand, et al., 1994, Methods in Cell Biology 44:635-654; Hay, et al., 1997, Proc. Natl. Acad. Sci. USA 94(10):5195-200). cDNAs or genomic regions encoding a normal or mutant insulin-like

20   gene can be operably fused to a desired promoter, as described above, and the promoter-insulin-like gene fusion inserted into any appropriate *Drosophila* transformation vector for the generation of transgenic flies. Typically, such transformation vectors are based on a well-characterized transposable elements, for example the P element (Rubin and Spradling, 1982, Science 218:348-53), the hobo element (Blackman, et al., 1989,

25   Embo J. 8(1):211-7), mariner element (Lidholm, et al., 1993, Genetics 134(3):859-68), the hermes element (O'Brochta, et al., 1996, Genetics 142(3):907-14), Minos (Loukeris, et al., 1995, Proc. Natl. Acad. Sci. USA 92(21):9485-9), or the PiggyBac element (Handler, et al., 1998, Proc. Natl. Acad. Sci. USA 95(13):7520-5), where the terminal repeat sequences of the transposon that are required for transposition are incorporated into the

30   transformation vector and arranged such that the terminal repeat sequences flank the transgene of interest (in this case a promoter-insulin-like gene fusion) as well as a marker gene used to identify transgenic animals. Most often, marker genes are used that affect the eye color of *Drosophila*, such as derivatives of the *Drosophila white* or *rosy* genes;

34

however, in principle, any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals, and examples of other marker genes used for transformation include the Adh^* gene used as a selectable marker for the transformation of Adh^- strains, $Ddc+$ gene used to transform Ddc^{ts2} mutant strains, the $lacZ$

5   gene of $E. coli$, and the $neomycin^R$ gene from the $E. coli$ transposon Tn5. Plasmid constructs for introduction of the desired transgene are coinjected into $Drosophila$ embryos having an appropriate genetic background, along with a helper plasmid that expresses the specific transposase need to mobilized the transgene into the genomic DNA. Animals arising from the injected embryos (G0 adults) are selected, or screened manually,

10   for transgenic mosaic animals based on expression of the marker gene phenotype and are subsequently crossed to generate fully transgenic animals (G1 and subsequent generations) that will stably carry one or more copies of the transgene of interest. Such stable transgenic animals are inspected for mutant phenotypes, such as abnormal development, morphology, metabolism, growth, longevity, reproduction, viability, or behavior, in order

15   to determine a function for the insulin-like gene created by ectopic expression or overexpression of the insulin-like gene, or by expression of a mutant insulin-like gene.

Generation of an overexpression/mis-expression phenotype is likely to result from either activation or inhibition of a receptor-linked signaling pathway. If such an overexpression/mis-expression phenotype is defined for an insulin-like gene, clonal

20   analysis can then be used to determine whether this phenotype is restricted to cells expressing the insulin-like gene (i.e. whether the phenotype is cell autonomous or cell non-autonomous). Methods of mitotic recombination of chromosomes in heterozygous flies can be used to generate mitotic clones of genetically homozygous cells that are well known to those skilled in the art, which include the use of X-rays or preferably FLP/FRT

25   mediated recombination (Xu and Harrison, 1994, Methods in Cell Biology 44:655-681; Greenspan, 1979, In $Fly Pushing: The Theory and Practice of Drosophila Genetics.$ Plainview, NY, Cold Spring Harbor Laboratory Press: pp. 103-124). These mitotic recombination techniques result in patches of cells, mitotic clones, that contain 2 or no copies of the gene-of-interest. Production of the overexpression/mis-expression phenotype

30   within cells in a clone having no copies of the gene-of-interest indicates that the effect is not cell autonomous, and is therefore likely to be the effect of a secreted molecule, as might be expected for insulin-like molecules.

35

Identification of Molecules that Interact With Insulin-Like Proteins

A variety of methods can be used to identify or screen for molecules, such as proteins or other molecules, that interact with insulin-like protein, or derivatives or fragments thereof. The assays may employ purified insulin-like protein, or cell lines or

5   model organisms such as *Drosophila* and *C. elegans*, that have been genetically engineered to express insulin-like protein. Suitable screening methodologies are well known in the art to test for proteins and other molecules that interact with insulin-like gene and protein (see *e.g.*, PCT International Publication No. WO 96/34099). The newly identified interacting molecules may provide new targets for pharmaceutical or

10 pesticidal agents. Any of a variety of exogenous molecules, both naturally occurring and/or synthetic (*e.g.*, libraries of small molecules or peptides, or phage display libraries), may be screened for binding capacity. In a typical binding experiment, the insulin-like protein or fragment is mixed with candidate molecules under conditions conducive to binding, sufficient time is allowed for any binding to occur, and assays are

15 performed to test for bound complexes. Assays to find interacting proteins can be performed by any method known in the art, for example, immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (*e.g.* by denaturing or nondenaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel

20 electrophoresis, *etc.*

Two-hybrid assay systems

A preferred method for identifying interacting proteins is a two-hybrid assay system or variation thereof (Fields and Song, Nature (1989) 340:245-246; U.S. Pat. No.

25 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1997) 31:663-704). The most commonly used two-hybrid screen system is performed using yeast. All systems share three elements: 1) a gene that directs the synthesis of a "bait" protein fused to a DNA binding domain; 2) one or more "reporter" genes having an upstream binding site for the bait, and 3) a gene that directs the synthesis of a "prey" protein fused

30 to an activation domain that activates transcription of the reporter gene. For the screening of proteins that interact with insulin-like protein, the "bait" is preferably a insulin-like protein, expressed as a fusion protein to a DNA binding domain; and the "prey" protein is a protein to be tested for ability to interact with the bait, and is

36

expressed as a fusion protein to a transcription activation domain. The prey proteins can be obtained from recombinant biological libraries expressing random peptides.

The bait fusion protein can be constructed using any suitable DNA binding domain, such as the *E. coli* LexA repressor protein, or the yeast GAL4 protein (Bartel *et al.*, BioTechniques (1993) 14:920-924, Chasman *et al.*, Mol. Cell. Biol. (1989) 9:4746-4749; Ma *et al.*, Cell (1987) 48:847-853; Ptashne *et al.*, Nature (1990) 346:329-331).

The prey fusion protein can be constructed using any suitable activation domain such as GAL4, VP-16, *etc.* The preys may contain useful moieties such as nuclear localization signals (Ylikomi *et al.*, EMBO J. (1992) 11:3681-3694; Dingwall and Laskey, Trends Biochem. Sci. Trends Biochem. Sci. (1991) 16:479-481) or epitope tags (Allen *et al.*, Trends Biochem. Sci. Trends Biochem. Sci. (1995) 20:511-516) to facilitate isolation of the encoded proteins.

Any reporter gene can be used that has a detectable phenotype such as reporter genes that allow cells expressing them to be selected by growth on appropriate medium (*e.g.* HIS3, LEU2 described by Chien *et al.*, PNAS (1991) 88:9572-9582; and Gyuris *et al.*, Cell (1993) 75:791-803). Other reporter genes, such as LacZ and GFP, allow cells expressing them to be visually screened (Chien *et al.*, *supra*).

Although the preferred host for two-hybrid screening is the yeast, the host cell in which the interaction assay and transcription of the reporter gene occurs can be any cell, such as mammalian (*e.g.* monkey, mouse, rat, human, bovine), chicken, bacterial, or insect cells. Various vectors and host strains for expression of the two fusion protein populations in yeast can be used (U.S. Pat. No. 5,468,614; Bartel *et al.*, Cellular Interactions in Development (1993) Hartley, ed., Practical Approach Series xviii, IRL Press at Oxford University Press, New York, NY, pp. 153-179; and Fields and Sternglanz, Trends In Genetics (1994) 10:286-292). As an example of a mammalian system, interaction of activation tagged VP16 derivatives with a GAL4-derived bait drives expression of reporters that direct the synthesis of hygromycin B phosphotransferase, chloramphenicol acetyltransferase, or CD4 cell surface antigen (Fearon *et al.*, PNAS (1992) 89:7958-7962). As another example, interaction of VP16-tagged derivatives with GAL4-derived baits drives the synthesis of SV40 T antigen, which in turn promotes the replication of the prey plasmid, which carries an SV40 origin (Vasavada *et al.*, PNAS (1991) 88:10686-10690).

37

Typically, the bait insulin-like gene and the prey library or chimeric genes are combined by mating the two yeast strains on solid or liquid media for a period of approximately 6-8 hours. The resulting diploids contain both kinds of chimeric genes, i.e., the DNA-binding domain fusion and the activation domain fusion.

5   Transcription of the reporter gene can be detected by a linked replication assay in the case of SV40 T antigen (described by Vasavada *et al.*, *supra*) or using immunoassay methods, preferably as described in Alam and Cook (Anal. Biochem. (1990)188:245-254). The activation of other reporter genes like URA3, HIS3, LYS2, or LEU2 enables the cells to grow in the absence of uracil, histidine, lysine, or leucine,

10   respectively, and hence serves as a selectable marker. Other types of reporters are monitored by measuring a detectable signal. For example, GFP and lacZ have gene products that are fluorescent and chromogenic, respectively.

After interacting proteins have been identified, the DNA sequences encoding the proteins can be isolated. In one method, the activation domain sequences or DNA-

15   binding domain sequences (depending on the prey hybrid used) are amplified, for example, by PCR using pairs of oligonucleotide primers specific for the coding region of the DNA binding domain or activation domain. Other known amplification methods can be used, such as ligase chain reaction, use of Q replicase, or various other methods described (see Kricka *et al.*, Molecular Probing, Blotting, and Sequencing (1995)

20   Academic Press, New York, Chapter 1 and Table IX). If a shuttle (yeast to *E. coli*) vector is used to express the fusion proteins, the DNA sequences encoding the proteins can be isolated by transformation of *E. coli* using the yeast DNA and recovering the plasmids from *E. coli*. Alternatively, the yeast vector can be isolated, and the insert encoding the fusion protein subcloned into a bacterial expression vector, for growth of

25   the plasmid in *E. coli*.

Immunoassays

Immunoassays can be used to identify proteins that interact with or bind to insulin-like protein. Various assays are available for testing the ability of a protein to

30   bind to or compete with binding to a wild-type insulin-like protein or for binding to an anti-insulin-like protein antibody. Suitable assays include radioimmunoassays, ELISA (enzyme linked immunosorbent assay), immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using

38

colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, immunoelectrophoresis assays, etc.

5

Biochemical Assays Using Insulin-Like Proteins

The present invention provides for biochemical assays using the insulin-like proteins. In one embodiment, Drosophila insulin-like proteins are useful for biochemical assays aimed at the identification and characterization of the ligand(s) for the known

10   Drosophila insulin receptor encoded by the InR (DIR) gene (Nishida, et al., 1986, Biochem. Biophys. Res. Commun. 141(2):474-81; Petruzzelli, et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83(13):4710-4; Fernandez-Almonacid and Rosen, 1987, Mol. Cell Biol. 7(8):2718-27), or the identification of ligands for new insulin-like receptor proteins that are discovered. The cDNAs encoding the insulin-like proteins can be individually

15   subcloned into any of a large variety of eukaryotic expression vectors permitting expression in insect and mammalian cells, described above. The resulting genetically engineered cell lines expressing insulin-like proteins can be assayed for production, processing, and secretion of the mature insulin-like proteins, which lack the secretory signal peptide and connecting C peptide regions, for example with antibodies to

20   Drosophila insulin-like proteins and Western blotting assays or ELISA assays. For assays of specific receptor binding and functional activation of receptor proteins, one can employ either crude culture medium or extracts containing secreted protein from genetically engineered cells (devoid of other insulin proteins), or partially purified culture medium or extracts, or preferably highly purified Drosophila insulin-like protein fractionated, for

25   example, by chromatographic methods. Alternatively, mature Drosophila insulin-like protein can be synthesized using chemical methods (Nagata, et al., 1992, peptides 13(4):653-62).

Specific protein binding of Drosophila insulin-like proteins to the Drosophila InR receptor can be assayed as follows, for example, following the procedures of Yamaguchi

30   et al. (Yamaguchi et al., 1995, Biochemistry 34:4962-4968). Chinese hamster ovary cells, COS cells, or any other suitable cell line, can be transiently transfected or stably transformed with expression constructs that direct the production of the Drosophila insulin receptor InR. Direct binding of a Drosophila insulin-like protein to such InR-expressing

cells can be measured using a "labeled" purified *Drosophila* insulin-like protein
derivative, where the label is typically a chemical or protein moiety covalently attached to
the insulin-like polypeptide which permits the experimental monitoring and quantitation of
the labeled *Drosophila* insulin-like protein in a complex mixture.

5   Specifically, the label attached to the insulin-like protein can be a radioactive
substituent such as an ^{125}I-moiety or ^{32}P-phosphate moiety, a fluorescent chemical
moiety, or labels which allow for indirect methods of detection such as a biotin-moiety
for binding by avidin or streptavidin, an epitope-tag such as a Myc- or FLAG-tag, or a
protein fusion domain which allows for direct or indirect enzymatic detection such as an
10   alkaline phosphatase-fusion or Fc-fusion domain. Such labeled *Drosophila* insulin-like
proteins can be used to test for direct and specific binding to InR-expressing cells by
incubating the labeled *Drosophila* insulin-like protein with the InR-expressing cells in
serum-free medium, washing the cells with ice-cold phosphate buffered saline to
remove unbound insulin-like protein, lysing the cells in buffer with an appropriate
15   detergent, and measuring label in the lysates to determine the amount of bound
insulin-like protein. Alternatively, in place of whole cells, membrane fractions obtained
from InR-expressing cells may also be used. Also, instead of a direct binding assay, a
competition binding assay may be used. For example, crude extracts or purified
Drosophila insulin-like protein can be used as a competitor for binding of labeled
20   purified bovine or porcine insulin to InR-expressing cells, by adding increasing
concentrations of *Drosophila* insulin-like protein to the mixture. The specificity and
affinity of binding of *Drosophila* insulin-like proteins can be judged by comparison
with other insulin superfamily proteins tested in the same assay, for example vertebrate
insulin, vertebrate IGF-I, vertebrate IGF-II, vertebrate relaxin, or silkmoth bombyxin.

25

Identification Of Additional Receptors Or Insulin-Like Binding Proteins

          The invention described herein provides for methods in which Drosophila
insulin-like proteins are used for the identification of novel insulin receptor proteins,
other than *Drosophila* InR, using biochemical methods well known to those skilled in
30   the art for detecting specific protein-protein interactions (Current Protocols in Protein
Science, 1998, Coligan et al., eds., John Wiley & Sons, Inc., Somerset, New Jersey).
Given the sequence diversity of the *Drosophila* insulin-like proteins detailed herein, the
identification to date of only a single insulin receptor gene in *Drosophila*, *InR*, points to

                                              40

the possibility that some *Drosophila* insulin-like proteins may bind to other receptors. In particular, it is possible that some *Drosophila* insulin-like proteins interact with receptor types that have not yet been discovered in vertebrates, for example the relaxin receptor, or receptor types that are specific to invertebrates. The identification of either

5   novel receptor types or invertebrate-specific receptor types is of great interest with respect to human therapeutic applications, or pesticide applications, respectively. Assuming some *Drosophila* insulin-like proteins do not exhibit specific protein binding to the known InR protein in the binding assays described above, then the novel cognate receptors for these insulin-like proteins can be investigated and identified as follows.

10   Labeled *Drosophila* insulin-like proteins can be used for binding assays *in situ* to identify tissues and cells possessing cognate receptors, for example as described elsewhere (Gorczyca et al., 1993, J. Neurosci. 13:3692-3704). Also, labeled *Drosophila* insulin-like proteins can be used to identify specific binding proteins including receptor proteins by affinity chromatography of *Drosophila* protein extracts using resins, beads,

15   or chips with bound *Drosophila* insulin-like protein (Formosa, et al., 1991, Methods Enzymol 208:24-45; Formosa, et al., 1983, Proc. Natl. Acad. Sci. USA 80(9):2442-6). Further, specific insulin-binding proteins can be identified by cross-linking of radioactively-labeled or epitope-tagged insulin-like protein to specific binding proteins in lysates, followed by electrophoresis to identify and isolate the cross-linked protein

20   species (Ransone, 1995, Methods Enzymol 254:491-7). Still further, molecular cloning methods can be used to identify novel receptors and binding proteins for *Drosophila* insulin-like proteins including expression cloning of specific receptors using *Drosophila* cDNA expression libraries transfected into mammalian cells, expression cloning of specific binding proteins using *Drosophila* cDNA libraries expressed in *E.*

25   *coli* (Cheng and Flanagan, 1994, Cell 79(1):157-68), and yeast two-hybrid methods (as described above) using a *Drosophila* insulin-like protein fusion as a "bait" for screening activation-domain fusion libraries derived from *Drosophila* cDNA (Young and Davis, 1983, Science 222(4625):778-82; Young and Davis, 1983, Proc. Natl. Acad. Sci. USA 80(5):1194-8; Sikela and Hahn, 1987, Proc. Natl. Acad. Sci. USA 84(9):3038-42;

30   Takemoto, et al., 1997, DNA Cell Biol 16(6):797-9).

ASSAYS OF INSULIN-LIKE PROTEINS

The functional activity of insulin-like proteins, derivatives and analogs can be

41

assayed by various methods known to one skilled in the art.

For example, in one embodiment, where one is assaying for the ability to bind to or compete with a wild-type insulin-like protein for binding to an anti-insulin-like protein antibody, various immunoassays known in the art can be used, including competitive and

5   non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays),

10   complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, *etc.* In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the

15   art for detecting binding in an immunoassay and are within the scope of the present invention. In another embodiment, where an insulin-like-binding protein is identified, the binding can be assayed, *e.g.*, by means well-known in the art. In another embodiment, physiological correlates of insulin-like protein binding to its substrates and/or receptors (*e.g.*, signal transduction) can be assayed.

20   In another embodiment, in insect (*e.g.*, Sf9 cells), fly (*e.g.*, *D. melanogaster*), or other model systems, genetic studies can be done to study the phenotypic effect of an insulin-like gene mutant that is a derivative or analog of a wild-type insulin-like gene. Other such methods will be readily apparent to the skilled artisan and are within the scope of the invention.

25

Other Functional Assays

For functional assays of *Drosophila* insulin-like protein, beyond receptor binding, the following activities can be investigated using InR-expressing cells after exposing said cells to crude or purified fractions of *Drosophila* insulin-like protein and comparing these

30   results with those obtained with other insulin superfamily proteins described above (Yamaguchi et al., 1995, Biochemistry 34:4962-4968). Assayable functional activities include stimulation of cell proliferation; stimulation of overall tyrosine kinase activity by immunoblotting of cell extracts with an anti-phosphotyrosine antibody; stimulation of

42

phosphorylation of specific substrate proteins such as InR or IRs-1 using ^{32}P- labeling and immunoprecipitation with antibodies that specifically recognize the substrate protein; and stimulation of other enzymatic activities linked to the insulin signaling pathway including assays of MAP kinase, Mek kinase, Akt kinase, and PI3-kinase activities.

5

Identifying Signaling Pathways And Phenotypes

This invention provides animal models which may be used in the identification and characterization of *D. melanogaster* insulin-like protein signaling pathways, and/or phenotypes associated with the mutation or abnormal expression of a *D. melanogaster*
10 insulin-like protein. Methods of producing such animal models using novel genes and proteins are well known in the art (*see e.g.*, PCT International Publication No. WO 96/34099, published October 31, 1996, which is incorporated by reference herein in its entirety). Such models include but are not limited to the following embodiments. Additional specific examples of animal models and their use are described below.

15 First, animals are provided in which a normal *D. melanogaster* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. Animals are also provided in which a normal gene has been recombinantly substituted for one or both copies of the animal's
20 homologous gene by homologous recombination or gene targeting.

Second, animals are provided in which a mutant *D. melanogaster* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. Animals are also provided in which a
25 mutant gene has been recombinantly substituted for one or both copies of the animal's homologous gene by homologous recombination or gene targeting.

Third, animals are provided in which a mutant version of one of that animal's own genes (bearing, for example, a specific mutation corresponding to, or similar to, a pathogenic mutation of an insulin-like gene from another species) has been recombinantly
30 introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment.

Finally, equivalents of transgenic animals, including animals with mutated or

43

inactivated genes, may be produced using chemical or x-ray mutagenesis. Using the isolated nucleic acids disclosed or otherwise enabled herein, one of ordinary skill may more rapidly screen the resulting offspring by, for example, direct sequencing, restriction fragment length polymorphism (RFLP) analysis, PCR, or hybridization analysis to detect

5   mutants, or Southern blotting to demonstrate loss of one allele.

Such animal models may be used to identify a *D. melanogaster* insulin-like protein signaling pathway by various methods. In one embodiment, this invention provides a method of identifying a *D. melanogaster* insulin-like protein signaling pathway comprising: (a) disrupting a *D. melanogaster* insulin-like gene; and (b) identifying the

10 effect of the gene disrupted in step (a) in an assay selected from the group consisting of a developmental assay, an energy metabolism assay, a growth rate assay and a reproductive capacity assay, lethality, sterility, reduced brood size, increased brood size, altered life span, defective locomotion, altered body shape, altered body plan, altered body size, altered bristles, altered body weight, altered cell size, increased cell division, decreased

15 cell division, altered feeding, slowed development, increased development, decreased metabolism (including alterations in glycogen synthesis, storage, and/or degradation, alterations in lipid synthesis, storage and/or degradation, alterations in levels of carbohydrate in hemolymph, alterations in levels of lipid in hemolymph), alterations in morphogenesis (including organs or tissues of the gonad, nervous system, fat body,

20 hemacytes, peripheral sensory organs, imaginal discs, eye, wing, leg, antennae, bristle, gut or musculature). Such assays are well known to those skilled in the art. In one embodiment, results of the assay may be compared to known mutant phenotypes to determine the signaling pathway involved. In one embodiment, the gene is disrupted using chemical mutagenesis. In another embodiment, the gene is disrupted using transposon

25 mutagenesis. In a further embodiment, the gene is disrupted by radiation mutagenesis.

Further, this invention provides a method of identifying a phenotype associated with mutation or abnormal expression of a *D. melanogaster* insulin-like protein comprising identifying the effect of a mutated or abnormally expressed *D. melanogaster* insulin-like gene in a *D. melanogaster* animal. In one embodiment, the effect is

30 determined by any of the assays mentioned above in connection with identifying a *D. melanogaster* insulin-like protein signaling pathway. The gene may be mutated or abnormally expressed using any technique known in the art, such chemical mutagenesis, radiation mutagenesis, transposon mutagenesis, antisense and double-stranded RNA

44

interference. Abnormal (i.e. ectopic) expression can be overexpression, underexpression (e.g., due to inactivation), expression at a developmental time different from wild-type animals, or expression in a cell type different from in wild-type animals.

Analysis Of Genetic Interactions And Multiple Mutants

5   Yet another approach that may be used to probe the biological function of the insulin-like gene identified herein is by using tests for genetic interactions with other genes that may participate in the same, related, interacting, or modifying genetic or biochemical pathways. Strains carrying mutations in multiple genes can be generated by cross breeding animals carrying the individual mutations, followed by selection of recombinant
10   progeny that carry the desired multiple mutations.

One specific question-of-interest is genetic analysis of interactions of insulin-like genes with other well-characterized *Drosophila* genes and pathways. Thus, double mutant fruit flies may be constructed that carry mutations in an insulin-like gene and another gene-of-interest.

15   It is of particular interest to test the interaction of the insulin-like gene with other genes implicated in insulin signaling, especially those that exhibit homology to insulin signaling components in vertebrates. For example, fruit flies carrying mutations in an insulin-like gene and either a loss-of-function mutation of *InR, chico, Pi3K92, Akt1, 14-3-3z, csw, Lar, Pk61C, Glut3, Ide, shaggy, s6k, Ras85D, drk, Sos, rl, or Dsorl*
20   (FlyBase 1998, "FlyBase - A Drosophila Database", Nucleic Acids Research 26:85-88; http://flybase.bio.indiana.edu), would be of use in investigating the involvement of different insulin-like genes in the signaling pathway where these genes participate. Similarly, transgenic animals mis-expressing an insulin-like gene which further carry mutations in the above-mentioned genes are of also of interest. Other genetic interactions
25   may be tested based on the actual phenotypes observed for alterations of the insulin-like gene alone.

Genetic Modifier Screens

The initial characterization of phenotypes created by mutations in the insulin-like
30   gene is expected to lead to the identification of *Drosophila* strains that exhibit mutant phenotypes suitable for large scale genetic modifier screens aimed at discovering other components of the same pathway. The procedures inv lved in typical genetic modifier screens to define other components of a genetic/biochemical pathway are well known to

45

those skilled in the art and have been described elsewhere (Wolfner and Goldberg, 1994,
Methods in Cell Biology 44:33-80; Karim et al., 1996, Genetics 143:315-329). Such
genetic modifier screens are based on the identification of mutations in other genes that
modify an initial mutant phenotype, by isolating either suppressor mutations that return the

5   mutant phenotype toward normal, or enhancer mutations that make the initial mutant
phenotype more severe.

Standard Genetic Modifier Screens

        Genetic modifier screens differ depending upon the precise nature of the mutant

10   allele being modified. If the mutant allele is genetically recessive, as is commonly the
situation for a loss-of-function allele, then most typically males, or in some cases females,
which carry one copy of the mutant allele are exposed to an effective mutagen, such as
EMS, MMS, ENU, triethylamine, diepoxyalkanes, ICR-170, formaldehyde, X-rays,
gamma rays, or ultraviolet radiation. The mutagenized animals are crossed to animals of

15   the opposite sex that also carrying the mutant allele to be modified, and the resulting
progeny are scored for rare events that result in a suppressed or enhanced version of the
original mutant phenotype. In the case where the mutant allele being modified is
genetically dominant, as is commonly the situation for ectopically expressed genes, wild
type males are mutagenized and crossed to females carrying the mutant allele to be

20   modified. Any new mutations identified as modifiers (i.e. suppressors or enhancers) are
candidates for genes that participate in the same phenotype-generating pathway.

        In a pilot-scale genetic modifier screen, 10,000 or fewer mutagenized progeny are
inspected; in a moderate size screen, 10,000 to 50,000 mutagenized progeny are inspected;
and in a large scale screen, over 50,000 mutagenized progeny are inspected. Progeny

25   exhibiting either enhancement or suppression of the original phenotype are immediately
crossed to adults containing balancer chromosomes and used as founders of a stable
genetic line. In addition, progeny of the founder adult are retested under the original
screening conditions to ensure stability and reproducibility of the phenotype. Additional
secondary screens may be employed, as appropriate, to confirm the suitability of each new

30   modifier mutant line for further analysis. For example, newly identified modifier
mutations can be tested directly for interaction with other genes of interest known to be
involved or implicated in insulin signaling pathways (*InR, chico, Pi3K92, Atk1, 14-3-3z,
csw, Lar, Pk61C, Glut3, Ide, shaggy, s6k, Ras85D, drk, Sos, rl, Dsor1*, mutations in other

46

insulin-like genes, or other modifier genes obtained from different genetic screens of the insulin signaling pathway), using methods described above. Als , the new modifier mutations can be tested for interactions with genes in other pathways thought to be unrelated or distantly related to insulin signaling, such as genes in the *Notch* signaling

5   pathway. New modifier mutations that exhibit specific genetic interactions with other genes implicated in insulin signaling, but not interactions with genes in unrelated pathways, are of particular interest. Additionally, strains can be generated that carry the new modifier mutations of interest in the absence of the original insulin-like gene mutation (i.e. a strain wild type for the mutant allele being suppressed or enhanced) to determine

10 whether the new modifier mutation exhibits an intrinsic phenotype, independent of the mutation in the insulin-like gene, which might provide further clues as to the normal function of the new modifier gene.

Each newly-identified modifier mutation can be crossed to other modifier mutations identified in the same screen to place them into complementation groups, which

15 typically correspond to individual genes (Greenspan, 1997, In *Fly Pushing: The Theory and Practice of Drosophila Genetics*, Plainview, NY, Cold Spring Harbor Laboratory Press: pp. 23-46). Two modifier mutations are said to fall within the same complementation group if animals carrying both mutations in *trans* exhibit essentially the same phenotype as animals that are homozygous for each mutation individually.

20

Gain-Of-Function Modifier Screens

Although the genetic modifier screens described above are quite powerful and sensitive, some genes that participate in an insulin-like pathway may be missed in this approach, particularly if there is functional redundancy of those genes. This is because the

25 vast majority of the mutations generated in the standard mutagenesis methods described above will be loss-of-function mutations, whereas gain-of-function mutations that could reveal genes with functional redundancy will be relatively rare. Another method of genetic screening in *Drosophila* has been developed that focuses specifically on systematic gain-of- function genetic screens (Rorth, et al., 1998, Development 125:1049-1057). This

30 method is based on a modular mis-expression system utilizing components of the GAL4/UAS system (which were defined above). In this case a modified P element, termed an EP element, is genetically engineered to contain a GAL4-responsive UAS element and promoter, and this engineered transposon is used to randomly tag genes by

47

insertional mutagenesis (similar to the method of P mutagenesis described above). Thousands of transgenic *Drosophila* strains, termed EP lines, can thus be generated each containing a specific UAS-tagged gene. This approach takes advantage of a well-recognized insertional preference of P elements, where it has been found that P

5   elements have a strong tendency to insert at the 5'-ends of genes. Consequently, many of the genes that have been tagged by insertion of EP elements become operably fused to a GAL4-regulated promoter, and increased expression or mis-expression of the randomly tagged gene can be induced by crossing in a GAL4 driver gene (similar that described above).

10   Thus, systematic gain-of-function genetic screens for modifiers of phenotypes induced by mutation or mis-expression of an insulin-like gene can be performed as follows. A large battery of thousands of *Drosophila* EP lines can be crossed into a genetic background containing a mutant or mis-expressed insulin-like gene, and further containing an appropriate GAL4 driver transgene. The progeny of this cross can be inspected for

15   enhancement or suppression of the original phenotype induced by mutation/mis-expression of the insulin-like gene. Progeny that exhibit an enhanced or suppressed phenotype can be crossed further to verify the reproducibility and specificity of this genetic interaction with the insulin-like gene. EP insertions that demonstrate a specific genetic interaction with a mutant or mis-expressed insulin-like gene, have therefore physically tagged a new gene

20   that genetically interacts with the insulin-like. The new modifier gene can be identified and sequenced using PCR or hybridization screening methods that allow the isolation of the genomic DNA adjacent to the position of the EP element insertion.

Assays For Changes In Gene Expression

25   This invention provides assays for detecting changes in the expression of the *D. melanogaster* insulin-like gene and proteins. Assays for changes in gene expression are well known in the art (*see e.g.*, PCT Publication No. WO 96/34099, published October 31, 1996, which is incorporated by reference herein in its entirety). Such assays may be performed *in vitro* using transformed cell lines, immortalized cell lines, or recombinant

30   cell lines, or *in vivo* using animal models.

        In particular, the assays may detect the presence of increased or decreased expression of a *D. melanogaster* insulin-like gene or protein on the basis of increased or decreased mRNA expression (using, *e.g.*, nucleic acid probes), increased or decreased

48

levels of related protein products (using, e.g., the antibodies disclosed herein), or increased or decreased levels of expression of a marker gene (e.g., -galactosidase or luciferase) operably linked to a 5' regulatory region in a recombinant construct.

5   In yet another series of embodiments, various expression analysis techniques may be used to identify genes which are differentially expressed between two conditions, such as a cell line or animal expressing a normal *D. melanogaster* insulin-like gene compared to another cell line or animal expressing a mutant *D. melanogaster* insulin-like gene. Such techniques comprise any expression analysis technique known to one skilled in the art, including differential display, serial analysis of gene expression (SAGE), nucleic acid

10   array technology, subtractive hybridization, proteome analysis and mass-spectrometry of two-dimensional protein gels. In a specific embodiment, nucleic acid array technology (*i.e.*, gene chips) may be used to determine a global (*i.e.*, genome-wide) gene expression pattern in a normal *D. melanogaster* animal for comparison with an animal having a mutation in a *D. melanogaster* insulin-like gene.

15   To elaborate further, the various methods of gene expression profiling mentioned above can be used to identify other genes (or proteins) that may have a functional relation to (*e.g.*, may participate in a signaling pathway with) a *D. melanogaster* insulin-like gene. Gene identification of such other genes is made by detecting changes in their expression levels following mutation, *i.e.*, insertion, deletion or substitution in, or overexpression,

20   underexpression, mis-expression or knock-out, of a *D. melanogaster* insulin-like gene, as described herein. Expression profiling methods thus provide a powerful approach for analyzing the effects of mutation in a *D. melanogaster* insulin-like gene.

Insulin-Like Gene Regulatory Elements

25   This invention provides methods for using insulin-like gene regulatory DNA elements to identify tissues, cells, genes and factors that specifically control insulin-like protein production. In one embodiment, regulatory DNA elements, such as enhancers/promoters, from a *Drosophila* insulin-like gene are useful for identifying and manipulating specific cells and tissues that synthesize an insulin-like protein. Such

30   hormone secreting cells and tissues are of considerable interest since they are likely to have an important regulatory function within the animal in sensing and controlling growth, development, reproduction, and/or metabolism. Analyzing components that are specific to insulin-like protein secreting cells is likely to lead to an understanding of how to

manipulate these regulatory processes, either for therapeutic applications or pesticide

applications, as well as an understanding of how to diagnose dysfunction in these

processes. For example, it is of specific interest to investigate whether there are

neuroendocrine tissues in *Drosophila* that might have a function related to that of the

5   mammalian pancreas in sensing and controlling metabolic activity through the production

of an insulin-like protein. Regulatory DNA elements derived from an insulin-like gene

provide a means to mark and manipulate such cells, and further, identify regulatory genes

and proteins, as described below.

10   **Protein-DNA Binding Assays**

In a third embodiment, insulin-like gene regulatory DNA elements are also useful

in protein-DNA binding assays to identify gene regulatory proteins that control the

expression of an insulin-like gene. Such gene regulatory proteins can be detected using a

variety of methods that probe specific protein-DNA interactions well known to those

15   skilled in the art (Kingston, 1998, In *Current Protocols in Molecular Biology*, Ausubel et

al, John Wiley & Sons, Inc., sections 12.0.3-12.10) including *in vivo* footprinting assays

based on protection of DNA sequences from chemical and enzymatic modification within

living or permeabilized cells, *in vitro* footprinting assays based on protection of DNA

sequences from chemical or enzymatic modification using protein extracts nitrocellulose

20   filter-binding assays and gel electrophoresis mobility shift assays using radioactively

labeled regulatory DNA elements mixed with protein extracts. In particular, it is of

interest to identify those DNA binding proteins whose presence or absence is specific to

insulin-like protein expressing tissue, as judged by comparison of the DNA-binding assays

described above using cells/extracts from an insulin-like gene expressing tissue versus

25   other cells/extracts from tissues that do not express insulin-like genes. For example, a

DNA-binding activity that is specifically present in cells that normally express an

insulin-like protein might function as a transcriptional activator of the insulin-like gene;

conversely, a DNA-binding activity that is specifically absent in cells that normally

express an insulin-like protein might function as a transcriptional repressor of the

30   insulin-like gene. Having identified candidate insulin-like gene regulatory proteins using

the above DNA-binding assays, these regulatory proteins can themselves by purified using

a combination of conventional and DNA-affinity purification techniques. In this case, the

DNA-affinity resins/beads are generated by covalent attachment to the resin of a small

50

synthetic double stranded oligonucleotide corresponding to the recognition site of the
DNA binding activity, or a small DNA fragment corresponding to the recognition site of
the DNA binding activity, or a DNA segment containing tandemly iterated versions of the
recognition site of the DNA binding activity. Alternatively, molecular cloning strategies

5   can be used to identify proteins that specifically bind insulin-like gene regulatory DNA
elements. For example, a *Drosophila* cDNA library in an *E. coli* expression vector, such
as the lambda-gt11 vector, can be screened for *Drosophila* cDNAs that encode insulin-like
gene regulatory element DNA-binding activity by probing the library with a labeled DNA
fragment, or synthetic oligonucleotide, derived from the insulin-like gene regulatory DNA,

10   preferably using a DNA region where specific protein binding has already been
demonstrated with a protein-DNA binding assay described above (Singh et al., 1989,
Biotechniques 7:252-61). Similarly, the yeast "one-hybrid" system can be used as another
molecular cloning strategy (Li and Herskowitz, 1993, Science 262:1870-4; Luo, et al.,
1996, Biotechniques 20(4):564-8; Vidal, et al., 1996, Proc. Natl. Acad. Sci. U.S.A.

15   93(19):10315-20). In this case, the insulin-like gene regulatory DNA element is operably
fused as an upstream activating sequence (UAS) to one, or typically more, yeast reporter
genes such as the *lacZ* gene, the *URA3* gene, the *LEU2* gene, the *HIS3* gene, or the *LYS2*
gene, and the reporter gene fusion construct(s) inserted into an appropriate yeast host
strain. It is expected that in the engineered yeast host strain the reporter genes will not be

20   transcriptionally active, for lack of a transcriptional activator protein to bind the UAS
derived from the *Drosophila* insulin-like gene regulatory DNA. The engineered yeast host
strain can be transformed with a library of *Drosophila* cDNAs inserted in a yeast
activation domain fusion protein expression vector, e.g. pGAD, where the coding regions
of the *Drosophila* cDNA inserts are fused to a functional yeast activation domain coding

25   segment, such as those derived from the GAL4 or VP16 activators. Transformed yeast
cells that acquire *Drosophila* cDNAs that encode proteins that bind the *Drosophila*
insulin-like gene regulatory element can be identified based on the concerted activation the
reporter genes, either by genetic selection for prototrophy (e.g. *LEU2, HIS3*, or *LYS2*
reporters) or by screening with chromogenic substrates (*lacZ* reporter) by methods known

30   in the art.

**Use Of *Drosophila* Insulin-Like Proteins As A Media Supplement For Gr wth And
Maintenance Of Insect Cells In Culture**

51

Not all insect cells can be propagated effectively in available media and furthermore it is difficult and time consuming to wean cells onto serum-free media for large scale protein production. The present invention provides for the use of *Drosophila* homologs of insulin-like proteins, as media additive for growth and maintenance of cells

5   in culture. Moreover, given that the *Drosophila* insulin-like proteins are the authentic endogenous protein hormones for *Drosophila* cells, and are likely to be more structurally and functionally similar to the authentic endogenous insulin-like hormones for other insect species, it is expected that *Drosophila* insulin-like hormones will exhibit superior properties in promoting growth and differentiation of insect cells in culture compared to

10   the effects found for mammalian insulins on insect cells.

In a specific embodiment, the Drosophila insulin-like proteins are used for the in vitro cultivation of *Drosophila* or other insect cells. Insect cell lines are widely used for basic research on the cell and molecular biology of insects. Also, *Drosophila* and other insect cell lines have application as a preferred system for developing cell-based assays for

15   insecticide targets, particularly those that might be amenable to high throughput screening methods (US Patent No. 5,767,261; US Patent No. 5,487,986; US Patent No. 5,641,652; US Patent No. 5,593,862; US Patent No. 5,593,864; US Patent No. 5,550,049; US Patent No. 5,514,578).

In another embodiment, the *Drosophila* insulin-like proteins are employed for the

20   *in vitro* cultivation of *Drosophila* and other insect cell lines used as host cells for the economical production of recombinant proteins on laboratory, pilot, or commercial scales. Further, the *Drosophila* and other insect cell lines can be used as hosts for the large-scale growth *in vitro* of viruses or bacteria that can be used as commercial insect control agents

Although fetal calf serum has been traditionally used as a media additive for the

25   growth of insect cells in culture, it has a number of serious disadvantages. First, fetal calf serum is expensive, and is often used in large amounts at concentrations typically between 5% to 15%. Occasionally, fetal calf serum is not available commercially. Also, there are batch-to-batch variations in the activity of fetal calf serum in stimulating cell growth, and some batches have been found to be toxic to insect cells in culture. Thus, there is a need

30   for substitutes for fetal calf serum in growth media for insect cells in culture, and the use of *Drosophila* insulin-like proteins for this purpose is expected to help fulfill this need.

Accordingly, *Drosophila* insulin-like proteins described herein can be used as an additive to insect cell growth media at concentrations preferably ranging from 5 ng/L to

0.5 g/L, and as a substitute for either fetal calf serum or mammalian insulin, for the
following purposes (a) promoting the propagation of continuous insect cell lines from
primary cultures; (b) promoting the differentiation and maintenance of specific insect cell
types in culture such as nerve cells, muscle cells, or fat body cells; (c) promoting the
5   propagation of insect cell lines *in vitro* for use in cell-based pesticide screening assays; (d)
promoting the propagation of insect cell lines *in vitro* for use in large-scale production of
recombinant proteins, natural protein products, or other natural products; and (e)
promoting the propagation of insect cells for the large-scale production of viruses and
bacteria which use insect cells as a host.

10

Agricultural Uses Of A Drosophila Insulin-Like Gene

In another embodiment of the invention, a *Drosophila* insulin-like gene may be
used in controlling agriculturally important pest species. For example, the proteins
disclosed herein, or analogs or derivatives thereof, may have activity in modifying the
15   growth, feeding and/or reproduction of crop-damaging insects, or insect pests of farm
animals or of other animals. In general, effective pesticides exert a disabling activity on
the target pest such as lethality, sterility, paralysis, blocked development, or cessation of
feeding. Examples of such pests include egg, larval, juvenile and adult forms of flies,
mosquitos, fleas, moths, beetles, cicadia, grasshoppers, and crickets.

20   Tests for such activities can be any method known in the art. Pesticides
comprising the nucleic acids of the *Drosophila* insulin like proteins may be prepared in a
suitable vector for delivery to a plant or animal. Examples of such vectors include
Agrobacterium tumefaciens Ti plasmid-based vectors for the generation of transgenic
plants or recombinant cauliflower mosaic virus for the incoulation of plant cells or
25   retrovirus based vectors for the introduction of genes into vertebrate animals (Burns et al.,
1993, Proc. Natl. Acad. Sci. USA 90:8033-37); and vectors based on transposable
elements for the introduction of genes into insects. For example, transgenic insects can be
generated using a transgene comprising an insulin-like gene operably fused to an
appropriate inducible promoter. For example, a tTA-responsive promoter may be used in
30   order to direct expression of the insulin-like protein at an appropriate time in the life cycle
of the insect. In this way, one may test efficacy as an insecticide in, for example, the larval
phase of the life cycle (i.e. when feeding does the greatest damage to crops).

Further, recombinant or synthetic insulin-like proteins, analogs, or derivatives can

53

be assayed for insecticidal activity by injection of solutions of insulin-like proteins into the hemolymph of insect larvae (Blackburn, et al., 1998, Appl. Environ. Microbiol. 64(8):3036-41; Bowen and Ensign, 1998, Appl. Environ. Microbiol. 64(8):3029-35). Still further, transgenic plants that express insulin-like proteins can be tested for activity against

5   insect pests (Estruch, et al., 1997, Nat. Biotechnol. 15(2):137-41).

In a preferred embodiment, an insulin-like gene can be tested as an insect control agent in the form of a recombinant virus that directs the expression of an insulin-like gene in the target pest. Suitable recombinant virus systems for expression of proteins in infected insect cells include recombinant Semliki Forest virus (DiCiommo and Bremner,

10   1998, J. Biol. Chem. 273:18060-66), recombinant sindbis virus (Higgs et al., 1995, Insect Mol. Biol. 4:97-103; Seabaugh et al., 1998, Virology 243:99-112), recombinant pantropic retrovirus (Matsubara et al., 1996, Proc. Natl. Acad. Sci. USA 93:6181-85; Jordan et al., 1998, Insect Mol. Biol. 7:215-22), and most preferably recombinant baculovirus. The use of recombinant baculovirus has a number of specific advantages including host specificity,

15   environmental safety, the availability of easily manipulable vector systems, and the potential use of the recombinant virus directly as a pesticide without the need for purification or formulation of the insulin-like protein. Thus, recombinant baculoviruses that direct the expression of an insulin-like gene can be used for both testing the pesticidal activity of insulin-like proteins under controlled laboratory conditions, and as insect

20   control agents in the field. One disadvantage of wild type baculoviruses as insect control agents can be the amount of time between application of the virus and death of the target insect, typically one to two weeks. During this period, the insect larvae continue to feed and damage crops. Consequently, there is a need to develop improved baculovirus-derived insect control agents which result in a rapid cessation of feeding of

25   infected target insects. The well-known metabolic regulatory role of insulins in vertebrates raises the possibility that expression of insulin-like proteins from recombinant baculovirus in infected insects may have a desirable effect in controlling metabolism and limiting feeding of insect pests.

Mutational anaylsis of insulin-like genes may also be used in connection with the

30   control of agriculturally-important pests. In this regard, mutational analysis of genes encoding insulin-like hormones in *Drosophila* provides a rational approach to determine the precise biological function of this class of hormones in invertebrates. Further, mutational analysis provides a means to validate potential pesticide targets that are

54

constituents of these signaling pathways.

Drosophila insulin-like genes, proteins or derivatives thereof may be formulated with any carrier suitable for agricultural use, such as water, organic solvents and/or inorganic solvents. The pesticide composition may be in the form of a solid or liquid composition and may be prepared by fundamental formulation processes including dissolving, mixing, milling, granulating, and dispersing.

The present invention encompasses compositions containing a Drosophila insulin-like protein or gene in a mixture with agriculturally acceptable excipients known in the art, such as vehicles, carriers, binders, UV blockers, adhesives, hemecants, thickeners, dispersing agents, preservatives and insect attractants. Thus the compositions of the invention may, for example, be formulated as a solid comprising the active agent and a finely divided solid carrier. Alternatively, the active agent may be contained in liquid compositions including dispersions, emulsions and suspensions thereof. Any suitable final formulation may be used, including for example, granules, powder, bait pellets (a solid composition containing the active agent and an insect attractant or food substance), microcapsules, water dispersible granules, emulsions and emulsified concentrates.

Examples of adjuvant or carriers suitable for use with the present invention include water, organic solvent, inorganic solvent, talc, pyrophyllite, synthetic fine silica, attapugus clay, kieselguhr chalk, diatomaceous earth, lime, calcium carbonate, bontonite, fuller's earth, cottonseed hulls, wheat flour, soybean flour, pumice, tripoli, wood flour, walnut shell flour, redwood flour, and lignin.

The compositions of the present invention may also include conventional insecticidal agents and/or may be applied in conjunction with conventional insecticidal agents.

EXAMPLE

Identification Of A *D. Melanogaster* Insulin-Like Gene, dIns1

A systematic cDNA cloning strategy was developed to identify genes that act during early mesodermal development in the fruit fly *Drosophila melanogaster* (Casal and Leptin, 1996, "Identification of novel genes in *Drosophila* reveals the complex regulation of early gene activity in the mesoderm", Proc. Natl. Acad. Sci. USA 93: 10327-10332). In brief, a subtractive cDNA library was generated containing cDNA inserts highly enriched for genes specifically expressed in the mesoderm. Positive

55

clones arising from this screen where placed into 16 different classes that corresponded
to different genes. Representative clones from each class were chosen for DNA
sequencing, and it was determined that three of the 16 classes had inserts identical to
known *Drosophila* genes, *tinman*, *DFR1*, and *zfh-1*. The remaining 13 classes appeared

5   to correspond to novel *Drosophila* genes, tentatively designated as genes *A-M*. Nine of
these genes were characterized by assaying their expression patterns by *in situ*
hybridization to wild type and mutant Drosophila embryos. Five of these nine genes
exhibited complicated expression patterns that were not entirely consistent with the
simple hypothesis that two regulatory genes, *twist* and *snail*, serve as the primary

10   zygotic regulators of early gene expression in the primordium of the mesoderm. Gene *A*
was found to exhibit a particularly unconventional pattern of expression. Gene *A* was
found to be expressed in a broad ventral domain independent of both *twist* and *snail*.
Furthermore, unlike any other known ventrally expressed gene, gene *A* was found to be
repressed posteriorly by *tailless*.

15   The unconventional expression pattern of gene *A* raised interest in the structure
and function of the protein encoded by this gene. The initial DNA sequence of 118 bp
of the representative cDNA clone for gene A, termed clone "B3", did not reveal an open
reading frame to provide structural information regarding the protein product
(GenBank, accession no. W91831). Accordingly, clone B3 was used to isolate a full-

20   length cDNA derived from gene A. The insert cDNA in clone B3 was purified, labeled
with ^{32}P, and used as a hybridization probe to screen another cDNA library, which
contained large cDNA inserts derived from *Drosophila* 4-8 hour embryonic mRNA
inserted into the vector pNB4O. Ten positive clones were identified in this screen, and
the plasmid DNAs were purified and analyzed by restriction enzyme digestion. The

25   results of this analysis indicated that all ten clones were structurally related, and the
clone containing the largest insert was picked for DNA sequencing (Sequenase II DNA
sequencing reagent kit, United States Biochemical). The complete sequence of this
cDNA clone revealed an insert of approximately 802 bp with one major open reading
frame of 405 bp (see annotated sequence in FIG. 4). Shown in Figure 4 is the annotated

30   sequence of dInsl, which contains a signal sequence followed by a B peptide, C peptide,
and A peptide, as indicated by the heavy lines below the respective sequences.

The predicted protein product of the gene A open reading frame was analyzed by
computational methods to obtain insight into its function. Analysis for ProSite protein
motifs demonstrated a clear match to an "insulin family signature"

35   CCxxxCxxxx(L,I,V,M,F)xxxC, beginning at amino acid 119 in the gene A predicted
protein (Biaroch et al., 1997, Nucleic Acids Res. 25:217-210). Furthermore, when the

56

gene A protein sequence was used as a query for searches of protein sequence databases
with the homology matching programs BLAST (Altschul et al., 1990, J. Mol. Biol.,
215:403-410; Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402) and FASTA
(Pearson et al., 1988, Proc. Nat'l Acad. Sci. USA, 85:2444-2448), the results

5   demonstrated highly significant matches between the gene A protein and many known
members of the insulin superfamily, including human insulin, sheep IGF2, chicken
insulin, MIP-I of freshwater snail, locust LIRP, pig relaxin, silkmoth bombyxin II, and
many others. Homologous regions were found in the gene A protein with both the B
peptide and A peptide regions of other well-characterized insulin superfamily proteins.

10   In keeping with the notion that gene A encodes a secreted hormone of the insulin
superfamily, analysis of the gene A predicted protein sequence with the PSORT
program (Nakai and Kanehisa, 1992, Genomics 14: 897-911) revealed the existence of a
secretory signal at the N-terminus with a site of predicted signal peptidase cleavage
between residues 20 and 21 of the precursor protein (see FIG. 4). Consequently, we

15   have renamed gene A as dIns1 (for *Drosophila* insulin-like gene 1) to acknowledge its
sequence similarity with other insulin superfamily genes.

To provide further confirmation that dIns1 encodes an insulin-like protein, the
ClustalW program was used to align the dIns1 protein sequence with other members of
the insulin superfamily, including locust LIRP, silkmoth bombyxin II, MIP-I from

20   freshwater snail, the F13B12 insulin-like protein from *C. elegans*, human insulin,
human IGF-I, and human relaxin (see FIG. 3, residues N-terminal to the B peptide
region or C-terminal to the A peptide region are not shown; residues in the large
cleavable C peptides are not shown except the dibasic residues at the prohormone
convertase cleavage sites). The alignment demonstrated that the dIns1 protein contains

25   all the key structural features known to be important for proper folding and processing
of insulin-like preprohormones. It is particularly notable that the dIns1 protein
possesses a large C peptide of more than 60 amino acids flanked by dibasic residues,
which are recognized by prohormone convertases during removal of the C peptide from
the prohormone. Also, the dIns1 precursor does not have a large C-terminal extension

30   beyond the A peptide, such as is found in the E peptide region of IGFs. Consequently,
the overall organization of the dIns1 protein is similar to that of vertebrate insulins or
relaxins, rather than that of vertebrate IGFs. This is of interest since it remains
uncertain when the structure and function of IGFs diverged from insulins during
metazoan evolution (McRory and Sherwood, 1997, DNA and Cell Biology 16: 939-49).

35   Importantly, the dIns1 protein possesses exactly the same number (six) and
spacing of Cys residues as found in vertebrate insulin proteins (boxed in FIG. 3),

57

indicating that the disulfide bonding pattern stabilizing the quaternary structure of dIns1 is likely to be identical to that of vertebrate insulins (see FIG. 1). This contrasts with the situation for some other invertebrate insulin-like proteins which have been found to have unusual disulfide features including an extra pair of Cys residues (illustrated in

5   FIG. 3 by MIP-I from freshwater snail, and F13B12 from the nematode *C. elegans*), lack of a conserved Cys pair (found in a *C. elegans* insulin-like subfamily, or altered spacing between Cys residues in the A or B chains (found in some *C. elegans* insulin-like proteins, (Smit et al., 1988, Nature 331:535-538; WO1999US08522; Duret, et al., 1998, Genome Res. 8(4):348-53; Kawano, et al., 1998, Worm Breeder's Gazette

10   15(2):47; Pierce and Ruvkun, 1998, Early 1998 East Coast Worm Meeting, abstract 150; Wisotzkey and Liu, 1998, Early 1998 East Coast Worm Meeting, abstract 206). It is also evident that the dIns1 protein has hydrophobic residues in positions that normally contribute to stabilizing the core structure at the interface between the A and B peptides in the folded protein (shaded in FIG. 3). Given the presence of these conserved

15   structural features within the dIns1 protein, it is highly likely that it will adopt a secondary and tertiary structure very similar to that found for the vertebrate and invertebrate insulin-like proteins that have been structurally analyzed to date. Specifically, expected features include a long central helix in the B peptide and two short antiparallel helices in the A peptide joined by a loop (see FIG. 3).

20

WHAT IS CLAIMED IS:

1. A purified protein comprising an amino acid sequence of an A peptide domain
and/or B peptide domain of dIns1 insulin-like gene, as depicted in Figure 3.

5   2. The purified protein of claim 1 comprising the amino acid sequence of SEQ ID
Nos 2.

3. A purified antibody, or derivative thereof containing the idiotype of the
antibody, capable of immunospecific binding to the protein of Claim 1 and not to an
10   insulin-like protein of another species.

4. An isolated nucleic acid comprising a nucleotide sequence encoding the amino
acid sequence depicted in Figure 4 (SEQ ID NO:2), wherein said nucleic acid is less than
15 kilobases.

15

5. The isolated nucleic acid of Claim 4 comprising the nucleotide sequence of
SEQ ID NO: 1.

6. A recombinant cell containing a recombinant nucleic acid vector comprising a
20   nucleotide sequence encoding an amino acid sequence as depicted in Figure 4 (SEQ ID
NO:2).

7. A method of identifying a phenotype associated with mutation or abnormal
expression of a *D. melanogaster* insulin-like protein comprising identifying an effect of a
25   mutated or abnormally expressed *D. melanogaster* insulin-like gene which encodes a *D.
melanogaster* insulin-like protein comprising an amino acid sequence as depicted in
Figure 4 (SEQ ID NO:2) in a *D. melanogaster* animal.

8. The method of Claim 7 wherein the gene is mutated or abnormally expressed
30   using a technique selected from the group consisting of radiation mutagenesis, chemical
mutagenesis, transposon mutagenesis, antisense and double-stranded RNA interference.

9. A modified, isolated *D. melanogaster* animal in which a *D. melanogaster*

59

insulin-like gene which encodes a *D. melanogaster* insulin-like protein comprising an amino acid sequence as depicted in Figure 4 (SEQ ID NO:2) which has been deleted or inactivated by recombinant methods, or a progeny thereof containing the deleted or inactivated gene.

5

10.   A recombinant non-human animal containing a *D. melanogaster* insulin-like transgene which encodes a *D. melanogaster* insulin-like protein comprising an amino acid sequence as depicted in Figure 4 (SEQ ID NO:2).

10   11.   A method of identifying a molecule that alters the expression level of a *D. melanogaster* insulin-like gene corresponding to a cDNA sequence as depicted in Figure 4 (SEQ ID NO:1), which method comprises:

(a) contacting a transgenic fly cell with one or more molecules, said transgenic fly cell having a transgene comprising a promoter or enhancer region of genomic DNA from 1

15   base to 6 kilobases upstream of the start codon of the cDNA sequence operably linked to a reporter gene; and

(b) determining whether the level of expression of the reporter gene is altered relative to the level of expression of the reporter gene in the absence of the one or more molecules.

20   12.   A cell culture medium or medium supplement comprising (a) a sterile liquid carrier, and (b) a protein or fragment thereof, functional in promoting cell growth, survival, or differentiation, said protein comprising at least 10 contiguous amino acids as depicted in Figure 4 (SEQ ID NO:2).

60
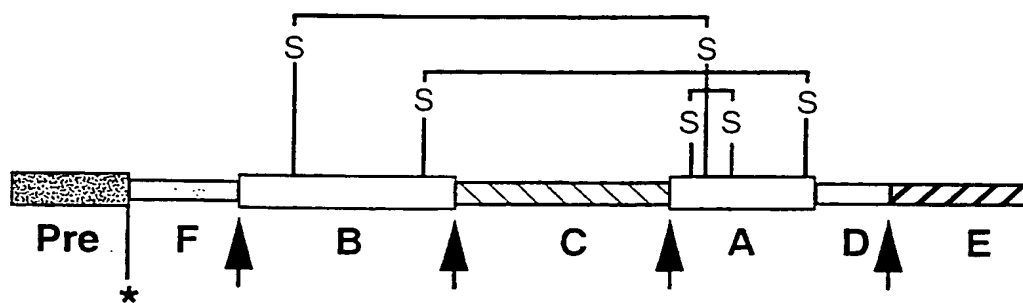
FIG._1

A peptide

```
        <--------->  <--------->       <---------->
GIVEQCCTSICSLYQLENYCN
...QTGIVDECCFRSCDLRRLEMYCAPLKPAKSA
RPYVALFEKCCLIGCTKRSLAKYC
AAATNPARHCCLSGCTRQDILTLCPH
SGRHRFDPFCCEVICDDGTSVKLCT
GIVDECCLRPCSVDVLLSYC
QGTTNLMECCMKPCTLSELRQMYC
TRGVFDECCRKTCSISELQTYYG

0       5      10      15      20
```

B peptide

```
helices  <--------->        <--------->
FVNQHLCGSHLVEALYLVCGERGFFYTPKT
GPETLCGAELVDALQFVCGDRGFYFNKPT....
KWKDDVIKLCGRELVRAQIAICGMSTWS
PAQEAPEKLCGHHFVRALVRLCGGPRWSPEDG
QLLRESLAAELRGCGPRFGKHLLSYCPMPEKTFTTTPGGWL
<QQPQAVHTYCGRHLARTLADLCWEAGVD
QFSACNINDRPHRRGVCSALADIVDFACSSSNQPAMV
SGAPQPVARYCGEKLSNALKLVCRGNYNTMF

0       5      10      15      20      25
```

human insulin
human IGF-I
human relaxin I
RLF
placentin
bombyxin II
MIP I
LIRP

FIG._2

FIG._3

```
        10          20          30          40          50          60
ATCAGTATCATTTGGCATGCCCAGCGATCGGTTTGCCAAGAGCACGAGAAGTTCGAGATA
TAGTCATAGTAAACCGTACGGGTCGCTAGCCAAACGGTTCTCGTGCTCTTCAAGCTCTAT


        70          80          90         100         110         120
GGACCCAGAGATACCAGAGATAAAGGAGGCATACCTTTTATGCCCGGTGAGAGCACGGAC
CCTGGGTCTCTATGGTCTCTATTTCCTCCGTATGGAAAATACGGGCCACTCTCGTGCCTG


       130         140         150         160         170         180
GGCGGAGTGAAAGATCGAGCAGGATGAGCCTGATTAGACTGGGACTGGCGCTGCTGCTCC
CCGCCTCACTTTCTAGCTCGTCCTACTCGGACTAATCTGACCCTGACCGCGACGACGAGG
                            M   S   L   I   R   L   G   L   A   L   L   L>
                            _____DINS1 CODING REGION_____>
                            M   S   L   I   R   L   G   L   A   L   L   L>
                                        _____SIGNAL PEPTIDE_____>


       190         200         210         220         230         240
TGCTGGCCACCGCGTCGCAGTTACTGCAGCCGGTCCAGGGACGCCGAAAGATGTGCGGCG
ACGACCGGTGGCGCAGCGTCAATGACGTCGGCCAGGTCCCTGCGGCTTTCTACACGCCGC
L   L   A   T   A   S   Q   L   L   Q   P   V   Q   G   R   R   K   M   C   G>
_____DINS1 CODING REGION_____>
L   L   A   T   A   S   Q   L>
_____SIGNAL PEPTIDE_____>
                    L   Q   P   V   Q   G   R   R   K   M   C   G>
                    _____B PEPTIDE_____>


       250         260         270         280         290         300
AGGCTCTGATCCAGGCACTGGATGTGATTTGTGTTAATGGATTTACACGCCGTGTCAGGC
TCCGAGACTAGGTCCGTGACCTACACTAAACACAATTACCTAAATGTGCGGCACAGTCCG
E   A   L   I   Q   A   L   D   V   I   C   V   N   G   F   T   R   R   V   R>
_____DINS1 CODING REGION_____>
E   A   L   I   Q   A   L   D   V   I   C   V   N   G   F   T>
_____B PEPTIDE_____>
                                        R   R   V   R>
                                        _____>


       310         320         330         340         350         360
GGAGCAGTGCGTCTAAGGATGCTAGAGTGCGAGACCTTATCCGTAAGCTACAGCAGCCGG
CCTCGTCACGCAGATTCCTACGATCTCACGCTCTGGAATAGGCATTCGATGTCGTCGGCC
R   S   S   A   S   K   D   A   R   V   R   D   L   I   R   K   L   Q   Q   P>
_____DINS1 CODING REGION_____>
R   S   S   A   S   K   D   A   R   V   R   D   L   I   R   K   L   Q   Q   P>
_____C PEPTIDE_____>


       370         380         390         400         410         420
ATGAGGACATTGAACAGGAAACGGAAACGGGAAGGTTAAAGCAGAAGCATACGGATGCGG
TACTCCTGTAACTTGTCCTTTGCCTTTGCCCTTCCAATTTCGTCTTCGTATGCCTACGCC
D   E   D   I   E   Q   E   T   E   T   G   R   L   K   Q   K   H   T   D   A>
_____DINS1 CODING REGION_____>
D   E   D   I   E   Q   E   T   E   T   G   R   L   K   Q   K   H   T   D   A>
_____C PEPTIDE_____>
```

FIG._4A-1

SUBSTITUTE SHEET (RULE 26)

```
              370         380         390         400         410         420
    ATGAGGACATTGAACAGGAAACGGAAACGGGAAGGTTAAAGCAGAAGCATACGGATGCGG
    TACTCCTGTAACTTGTCCTTTGCCTTTGCCCTTCCAATTTCGTCTTCGTATGCCTACGCC
     D   E   D   I   E   Q   E   T   E   T   G   R   L   K   Q   K   H   T   D   A>
    _____DINS1 CODING REGION_____>
     D   E   D   I   E   Q   E   T   E   T   G   R   L   K   Q   K   H   T   D   A>
    _____C PEPTIDE_____>

              430         440         450         460         470         480
    ATACGGAGAAGGGTGTGCCACCGGCCGTCGGAAGTGGACGAAAGTTGCGACGCCATCGGC
    TATGCCTCTTCCCACACGGTGGCCGGCAGCCTTCACCTGCTTTCAACGCTGCGGTAGCCG
     D   T   E   K   G   V   P   P   A   V   G   S   G   R   K   L   R   R   H   R>
    _____DINS1 CODING REGION_____>
     D   T   E   K   G   V   P   P   A   V   G   S   G   R   K   L   R   R   H   R>
    _____C PEPTIDE_____>

              490         500         510         520         530         540
    GACGCATCGCCCACGAGTGTTGCAAGGAGGGCTGCACCTACGACGATATACTGGACTACT
    CTGCGTAGCGGGTGCTCACAACGTTCCTCCCGACGTGGATGCTGCTATATGACCTGATGA
     R   R   I   A   H   E   C   C   K   E   G   C   T   Y   D   D   I   L   D   Y>
    _____DINS1 CODING REGION_____>
     R   R>
    _____>
         I   A   H   E   C   C   K   E   G   C   T   Y   D   D   I   L   D   Y>
    _____A PEPTIDE_____>

              550         560         570         580         590         600
    GCGCCTGATGACCAGGATGGCAAAACAAAACAAATAAAAACCAGAAACCAGATCCCAAAA
    CGCGGACTACTGGTCCTACCGTTTTGTTTTGTTTATTTTTGGTCTTTGGTCTAGGGTTTT
     C   A   *>
    _____>
     C   A>
    _____>
```

FIG._4A-2

6 / 6

```
        610       620       630       640       650       660
ACCAAGTACCAGATGAACACGACATGGCTGAGATTTTGTGTGGCGGCACGGGGAAAACAC
TGGTTCATGGTCTACTTGTGCTGTACCGACTCTAAAACACACCGCCGTGCCCCTTTTGTG


        670       680       690       700       710       720
CCGACGACCGGCAGGCTATTTGCAATTCATTTTCCTACTACACTTAACCCCTAACTATAA
GGCTGCTGGCCGTCCGATAAACGTTAAGTAAAAGGATGATGTGAATTGGGGATTGATATT


                                              >Poly(A)_signal
                                                    |
        730       740       750       760       770       780
ACGTAATCGTATTTCCAAATATTTCATTGTAAAATTTCTAGTGGAGGCAAATAAAGTTAC
TGCATTAGCATAAAGGTTTATAAAGTAACATTTTAAAGATCACCTCCGTTTATTTCAATG


        790       800
TCTCCAAGCAGCAGCAGAAAC
AGAGGTTCGTCGTCGTCTTTG
```

FIG._4B

SEQUENCE LISTING

<110> Leptin, Maria

<120> NUCLEIC ACIDS AND PROTEINS OF A D. MELANOGASTER
      INSULIN-LIKE GENE AND USES THEREOF

<130> 7326-077

<140>
<141>

<160> 3

<170> PatentIn Ver. 2.0

<210> 1
<211> 801
<212> DNA
<213> Drosophila melanogaster

<400> 1
```
atcagtatca tttggcatgc ccagcgatcg gtttgccaag agcacgagaa gttcgagata   60
ggacccagag ataccagaga taaaggaggc atacctttta tgcccggtga gagcacggac  120
ggcggagtga aagatcgagc aggatgagcc tgattagact gggactggcg ctgctgctcc  180
tgctggccac cgcgtcgcag ttactgcagc cggtccaggg acgccgaaag atgtgcggcg  240
aggctctgat ccaggcactg gatgtgattt gtgttaatgg atttacacgc cgtgtcaggc  300
ggagcagtgc gtctaaggat gctagagtgc gagaccttat ccgtaagcta cagcagccgg  360
atgaggacat tgaacaggaa acggaaacgg gaaggttaaa gcagaagcat acggatgcgg  420
atacggagaa gggtgtgcca ccggccgtcg gaagtggacg aaagttgcga cgccatcggc  480
gacgcatcgc ccacgagtgt tgcaaggagg gctgcaccta cgacgatata ctggactact  540
gcgcctgatg accaggatgg caaaacaaaa caaataaaaa ccagaaacca gatcccaaaa  600
accaagtacc agatgaacac gacatggctg agattttgtg tggcggcacg gggaaaacac  660
ccgacgaccg gcaggctatt tgcaattcat tttcctacta cacttaaccc ctaactataa  720
acgtaatcgt atttccaaat atttcattgt aaaatttcta gtggaggcaa ataaagttac  780
tctccaagca gcagcagaaa c                                            801
```

<210> 2
<211> 134
<212> PRT
<213> Drosophila melanogaster

<400> 2
```
Met Ser Leu Ile Arg Leu Gly Leu Ala Leu Leu Leu Leu Leu Ala Thr
 1               5                  10                  15


Ala Ser Gln Leu Leu Gln Pro Val Gln Gly Arg Arg Lys Met Cys Gly
            20                  25                  30
```

Glu Ala Leu Ile Gln Ala Leu Asp Val Ile Cys Val Asn Gly Phe Thr
        35   40   45

Arg Arg Val Arg Arg Ser Ser Ala Ser Lys Asp Ala Arg Val Arg Asp
        50   55   60

Leu Ile Arg Lys Leu Gln Gln Pro Asp Glu Asp Ile Glu Gln Glu Thr
65   70   75   80

Glu Thr Gly Arg Leu Lys Gln Lys His Thr Asp Ala Asp Thr Glu Lys
            85   90   95

Gly Val Pro Pro Ala Val Gly Ser Gly Arg Lys Leu Arg Arg His Arg
            100   105   110

Arg Arg Ile Ala His Glu Cys Cys Lys Glu Gly Cys Thr Tyr Asp Asp
        115   120   125

Ile Leu Asp Tyr Cys Ala
        130

<210> 3
<211> 15
<212> PRT
<213> Drosophila melanogaster

<220>
<221> SITE
<222> (3)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (4)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (5)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (7)
<223> Xaa=any amino acid

```
<220>
<221> SITE
<222> (8)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (9)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (10)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (11)
<223> Xaa=Leu, Ile, Val, Met, or Phe

<220>
<221> SITE
<222> (12)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (13)
<223> Xaa=any amino acid

<220>
<221> SITE
<222> (14)
<223> Xaa=any amino acid

<400> 3
Cys Cys Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys
  1               5                   10                  15
```

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 99/09316

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/17   C12N15/18   C12N5/14   C12Q1/02   C07K16/26
       A01K67/027 C07K14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12Q C12N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CASAL ET AL.: "Identification of novel genes in Drosophila reveals the complex regulation of early gene activity in the mesoderm" EMBL DATABASE ACC. NO: W91831, 9 November 1996 (1996-11-09), XP002134614 cited in the application	1-6
Y	the whole document -& CASAL ET AL.: "Identification of novel genes in Drosophila reveals the complex regulation of early gene activity in the mesoderm" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA, vol. 93, September 1996 (1996-09), pages 10327-10332, XP002134615 cited in the application figures 2,3; table 1	12

-/--

| X | Further documents are listed in the continuation of box C. | | Patent family members are listed in annex. |

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 April 2000	17/04/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	van Klompenburg, W

Form PCT/ISA/210 (second sheet) (July 1992)

page 1 of 2

BNSDOCID: <WO___0032779A1_I_>

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category ?	Citation of document. with indication,where appropriate. of the relevant passages	Relevant to claim No.
Y	PIMENTEL ET AL.: "Insulin acts as an embryonic growth factor for Drosophila neural cells" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 226, 1996, pages 855-861, XP002134616 cited in the application	12
A	page 855 page 857	1-11
X	HARVEY ET AL.: "BDGP/HHMI Drosophila EST Project" EMBL DATABASE ACC NO: DMAA40991, 2 January 1997 (1997-01-02), XP002134617 the whole document	1-6
A	MOSNA ET AL.: "Insulin can completely replace serum in Drosophila melanogaster cell cultures in vitro" EXPERIENTIA, vol. 37, 1981, pages 466-467, XP002134618 tables 1,2	1-12